L. Pandolfi

Lezioni di Analisi Matematica 2

Questa versione è scritta in carattere romano e giustificata a destra. Una versione più leggibile da chi ha problemi di dislessia (scritta in carattere bastone e non giustificata) ha per titolo *Lezioni di Analisi Matematica 2 (per studenti dislessici)*

Una versione audio per studenti ipo- o non vedenti è stata realizzata dal Laboratorio S. Polin dell'Università di Torino e può trovarsi all'indirizzo seguente:

`http://www.integr-abile.unito.it/knowledge-transfer/accessible-library-2/`

L. Pandolfi, in pensione dal Politecnico di Torino, Dipartimento di Scienze Matematiche "G.L. Lagrange"

Il testo presenta tre blocchi principali di argomenti:

A Successioni e serie numeriche e di funzioni: Cap. 1, e 2.

B Questa parte consta di due, da studiarsi in sequenza.
B1 Funzioni di più variabili e integrazione (multipla, di curva e di superficie): Cap. 3–8.
B2 Campi conservativi,Cap. 9.

C Sistemi di equazioni differenziali: Cap. 10.

Lo studio dei blocchi **A** e **B** può scambiarsi di ordine senza problemi. Invece, è consigliabile studiare **C** per ultimo. Infatti, lo studio del Cap. 10 richiede il concetto di continuità e differenziabilità di funzioni di più variabili, studiato ai paragrafi 4.1–4.2.

Ovunque nello studio del Cap. 10 è necessario conoscere il concetto di curva (ma non le proprietà differenziali delle curve, né gli integrali di curva).

L'esponenziale di matrici richiede la definizione di serie, Cap. 1, e 2 e il paragrafo 10.4 richiede il Cap. 9.

Indice

Capitolo 1

Serie numeriche

Le serie numeriche vogliono generalizzare la somma di un numero finito di termini al caso in cui si sommano infiniti termini. Per questo si introduce il limite di una opportuna successione di “somme parziali”. Prima di tutto quindi ricapitoleremo i concetti fondamentali relativi alle successioni numeriche.

1.1 Successioni numeriche: ricapitolazione

Una *successione numerica* è una funzione definita su $\mathbf{N}$ ed a valori in $\mathbf{R}$ (oppure in $\mathbf{C}$. Noi qui ci limitiamo a considerare successioni a valori reali).

Una successione si indica col simbolo (x_n) e si sottintende che $n \in \mathbf{N}$. Talvolta, n è un qualsiasi numero intero maggiore od uguale ad un certo n_0 che può anche essere negativo. Se è necessario specificare il primo dei valori dell'indice n scriveremo $(x_n)_{n \geq n_0}$.

La successione si chiama:

- “limitata” quando esiste M tale che $|x_n| < M$ per ogni n;
- “convergente”, quando esiste finito il limite $\lim_{n\to+\infty} x_n$, che spesso si indica semplicemente come $\lim x_n$;
- “divergente” quando $\lim x_n = +\infty$ oppure quando $\lim x_n = -\infty$;
- “regolare” quando è convergente oppure divergente;
- “oscillante” quando non è regolare; ossia quando $\lim x_n$ non esiste, né finito né $+\infty$ né $-\infty$.

Ricordiamo che una successione si dice *fondamentale* o *di Cauchy* quando

per ogni $\epsilon > 0$ esiste un N_ϵ tale che per ogni $n > N_\epsilon$ e per ogni $m > 0$ si ha:

$$|x_n - x_{n+m}| < \epsilon\,.$$

In simboli:

$$\forall \epsilon > 0\ \exists N_\epsilon \ \mid \forall n > N_\epsilon\,, \forall m > 0 \implies |x_n - x_{n+m}| < \epsilon\,.$$

Vale:

Teorema 1 *Ogni successione convergente è fondamentale; ogni successione fondamentale è limitata e quindi ogni successione convergente è limitata.*

Naturalmente, esistono successioni limitate e non convergenti. Per esempio la successione di termine generale $x_n = (-1)^n$. Invece:

Teorema 2 *Ogni successione fondamentale è convergente.*

Dim. Accenniamo ai passi cruciali della dimostrazione, che si trova nei testi di Analisi Matematica 1.

Sia (x_n) la successione. Prima di tutto si prova che la successione (x_n) ammette s.successioni convergenti. Questo si vede così: dato che una successione fondamentale è limitata, l'immagine di (x_n), ossia l'insieme $\{x_n\}$ è limitato. Se è finito, almeno uno dei suoi elementi è immagine di infiniti n e quindi la successione (x_n) ha una s.successione (x_{n_k}) costante e quindi convergente, di ciamo ad x_0.

Se l'insieme $\{x_n\}$ è infinito, esso ammette almeno un punto di accumulazione x_0, per il Teorema di Bolzano-Weierstrass. Si costruisce quindi una s.successione (x_{n_k}) convergente ad x_0.

L'ultimo passo della dimostrazione consiste nel mostrare che è la successione (x_n) stessa che converge ad x_0, usando la definizione di successione fondamentale. ∎

Una successione (x_n) è *crescente* quando $n > m$ implica $x_n \geq x_m$; *decrescente* quando $n > m$ implica $x_n \leq x_m$.

Un altro risultato importante da ricordare è il teorema delle funzioni monotone, la cui formulazione particolarizzata al caso delle successioni è la seguente:

Teorema 3 *Sia (x_n) una succesione monotona. Esiste* $\lim x_n$; *ossia,* ogni successione monotona è regolare.

Infine, ricordiamo che se una successione (x_n) è regolare, anche la successione che si ottiene da essa trascurandone un numero finito di termini, ossia $(x_n)_{n>m}$ è regolare, ed ha il medesimo limite[1].

[1] in realtà vale di più: ogni sottosuccessione (x_{n_k}) ha il medesimo limite della (x_n).

1.2 Le serie numeriche

Sia (x_n) una successione di numeri. Per fissare le idee sia $n \geq 1$, ma in modo analogo si può trattare il caso in cui il primo indice sia per esempio 0 o comunque sia diverso da 1.

Si chiama *serie* dei numeri x_n una nuova successione (s_n) costruita come segue:

$$s_1 = x_1 \,, \qquad s_2 = x_1 + x_2 \,, \qquad s_k = \sum_{n=1}^{k} x_n \,, \tag{1.1}$$

ossia, detto in modo più conciso:

$$s_1 = x_1 \,, \qquad s_k = s_{k-1} + x_k \,.$$

I numeri s_n si chiamano le *somme parziali* della serie[2]. La nuova successione (s_n) si indica anche col simbolo

$$\sum_{n=1}^{\infty} x_n \qquad \text{o, più semplicemente,} \qquad \sum x_n \,.$$

I numeri x_n si chiamano i *termini* della serie e si dice che x_n è il *termine generale* della serie.

Nella definizione precedente niente si richiede al comportamento della successione (x_n) o della successione (s_n). Se però la successione (s_n) converge allora si dice che *la serie converge*; se la successione (s_n) diverge (a $+\infty$ oppure a $-\infty$) allora si dice che *la serie diverge* (rispettivamente a $+\infty$ oppure a $-\infty$). Se la successione (s_n) è priva di limite, si dice che la serie è *oscillante* o *indeterminata*.

Una serie si dice *regolare* quando converge oppure diverge.

Il *carattere della serie* o *comportamento della serie* è la proprietà di essere convergente, divergente o oscilante.

Ricapitolando, se la successione (s_n) converge ad l oppure diverge, dovremmo indicare questo col simbolo

$$\lim_{k \to +\infty} \sum_{n=1}^{k} x_n = \alpha \,,$$

[2] La successione $\{s_n\}$ delle somme parziali si costruisce dalla (1.1), conoscendo la successione $\{x_n\}$. E' importante notare che, viceversa, nota la successione $\{s_n\}$, si può ricostruire la successione $\{x_n\}$. Infatti è

$$x_1 = s_1 \,, \text{ e, per } k > 1, \; x_k = s_k - s_{k-1} \,. \quad \blacksquare$$

α rispettivamente uguale a l oppure $+\infty$ oppure $-\infty$. Più brevemente si scrive

$$\sum_{n=1}^{\infty} x_n = \alpha \qquad \text{o anche} \quad \sum x_n = \alpha\,.$$

Si dice brevemente che α è la *somma* della serie.

Ovviamente, scambiando l'ordine di un numero finito di termini di una serie, non si cambia nè il comportamento della serie nè la sua somma, nel caso che la serie sia convergente (diremo più avanti cosa accade scambiando tra loro infiniti termini della serie). E' anche vero che, sopprimento o aggiungendo un numero finito di termini, oppure cambiando il valore di un numero finito di termini, la serie rimane convergente, divergente o oscillante; ossia:

Teorema 4 *Il carattere di una serie non muta alterandone un numero finito di termini.*

> Va detto esplicitamente che se la serie è convergente, la somma della serie cambia alterandone un numero finito di termini. Se invece è divergente, la sua somma non cambia.

Inoltre:

Teorema 5 *Se* $\sum x_n$ *converge allora* $\lim_{n\to+\infty} x_n = 0$.

Dim. Si indichi con $s_k = \sum_{n=1}^{k} x_n$. L'ipotesi è che la successione (s_k) converge e quindi anche la successione s_{k-1} converge, ed al medesimo limite. Dunque,

$$0 = \lim_{k\to+\infty} s_k - \lim_{k\to+\infty} s_{k-1} = \lim_{k\to+\infty} (s_k - s_{k-1}) = \lim_{k\to+\infty} x_k\,. \quad \blacksquare$$

Di conseguenza:

Esempio 6 La serie di "termine generale" $(-1)^n n$, ossia la serie

$$\sum (-1)^n n$$

non converge. ∎

Invece:

Esempio 7 La successione

$$(q^n)_{n\geq 0}$$

(con $q \in \mathbf{R}$ fissato) si chiama *progressione geometrica* (di *ragione* q). La serie

$$\sum_{n=0}^{+\infty} q^n$$

si chiama *serie geometrica*. E' noto che, se $q \neq 1$,

$$\sum_{n=0}^{N} q^n = \frac{1-q^{N+1}}{1-q}$$

e quindi

$$\lim_{N\to+\infty} \frac{1-q^{N+1}}{1-q} = \sum_{n=0}^{+\infty} q^n = \begin{cases} 1/(1-q) & \text{se} \quad |q|<1 \\ +\infty & \text{se} \quad q \geq 1 \\ \text{oscillante} & \quad \text{altrimenti.} \end{cases}$$

Si noti che la serie geometrica per definizione inizia con l'indice $n = 0$. Se per qualche ragione si deve iniziare con un primo indice diverso, di ciò va tenuto conto nel calcolo della somma. Per esempio

$$\sum_{n=0}^{+\infty} \frac{1}{2^n} = \frac{1}{1-1/2} = 2\,, \qquad \sum_{n=1}^{+\infty} \frac{1}{2^n} = 2-1 = 1\,. \quad \blacksquare$$

1.2.1 Serie telescopiche

Sia $(b_k)_{k\geq 0}$ una successione e sia

$$a_n = b_n - b_{n-1}$$

(ovviamente definita per $n \geq 1$). Consideriamo la serie

$$\sum_{n=1}^{+\infty} a_n\,. \tag{1.2}$$

Una serie ottenuta con questo procedimento si chiama *serie telescopica*.

E' facile calcolare le somme parziali di una serie telescopica:

$$s_1 = a_1 = b_1 - b_0\,, \quad s_2 = a_1 + a_2 = (b_1 - b_0) + (b_2 - b_1) = b_2 - b_0$$

e, in generale,

$$s_k = b_k - b_0\,.$$

Dunque:

Teorema 8 *La serie telescopica costruita sopra converge se e solo se*

$$\lim b_k = l \in \mathbf{R}$$

e in tal caso

$$\sum_{n=1}^{+\infty} a_n = l - b_0 \,;$$

diverge se $\lim b_k = +\infty$ *oppure se* $\lim b_k = -\infty$*. La serie è oscillante se e solo se la successione* (b_k) *è priva di limite.*

Esempi

Vediamo alcuni esempi.

Esempio 9 Consideriamo la serie

$$\sum_{n=1}^{+\infty} \log\left(1 + \frac{1}{n}\right) \,.$$

Questa serie **diverge.** Infatti,

$$\log\left(1 + \frac{1}{n}\right) = \log \frac{n+1}{n} = \log(n+1) - \log n \,.$$

Sia ha quindi una serie telescopica e

$$\sum_{n=1}^{k} \log\left(1 + \frac{1}{n}\right) = \log(k+1) \qquad \text{da cui} \quad \lim_{k\to+\infty} \sum_{n=1}^{k} \log\left(1 + \frac{1}{n}\right) = +\infty \,. \quad \blacksquare$$

Esempio 10 Consideriamo la serie

$$\sum_{n=1}^{+\infty} \frac{1}{n^2 + n} \,.$$

Si vede che questa è una serie telescopica notando che

$$\frac{1}{n^2+n} = \frac{1}{n} - \frac{1}{n+1} = -\left[\frac{1}{n+1} - \frac{1}{n}\right]$$

e inoltre $b_n = 1/n \to 0$. Dunque,

$$\sum_{n=1}^{+\infty} \frac{1}{n^2 + n} = 1 \,.$$

Se per qualche ragione si devono sommare i termini con $n \geq n_0$, allora

$$\sum_{n=n_0}^{+\infty} \frac{1}{n^2+n} = \frac{1}{n_0}\,.$$

Consideriamo ora la serie

$$\sum_{n=1}^{+\infty} \frac{1}{4n^2+8n+3}\,.$$

Decomponendo in fratti semplici,

$$\frac{1}{4n^2+8n+3} = \frac{-1}{4}\left[\frac{1}{n+3/2} - \frac{1}{n+1/2}\right] = \frac{1}{4}[b_{n+1} - b_n]\,, \qquad b_n = \frac{-1}{n+1/2}\,.$$

Si tratta quindi di una serie telescopica, la cui somma è 1/6. ∎

Infine:

Esempio 11 Anche la serie seguente è una serie telescopica:

$$\sum_{n=2}^{+\infty} \log\left[\frac{1}{\sqrt[n]{(n+1)^{1/(n-1)}}}\left(1+\frac{1}{n}\right)^{1/(n-1)}\right]\,.$$

Infatti,

$$\left(1+\frac{1}{n}\right)^{1/(n-1)} \frac{1}{\sqrt[n]{(n+1)^{1/(n-1)}}} = \frac{(n+1)^{1/n}}{n^{1/(n-1)}}$$

e quindi la serie è uguale a

$$\sum_{n=2}^{+\infty}\left[\frac{1}{n}\log(n+1) - \frac{1}{n-1}\log n\right] = \sum_{n=2}^{+\infty}[b_{n+1} - b_n]\;, \qquad b_n = \frac{1}{n}\log(n+1)\,.$$

Dunque, la serie converge e la sua somma è $-(1/2)\log 3$. ∎

Nonostante gli esempi importanti della serie geometrica e delle serie telescopiche, calcolare esplicitamente le somme parziali di una serie è pressoché impossibile. L'unica cosa che si può fare è dare condizioni per la convergenza o divergenza di serie, e quindi, se già si sa che la serie converge, approssimarne numericamente la somma.

1.3 Criteri di convergenza

Come si è detto, è ben difficile calcolare esplicitamente le somme parziali di una serie. Per questo è necessario conoscere dei criteri che assicurino la convergenza o meno di una serie, senza calcolarne le somme parziali. Dato che la somma di una serie è il limite della successione delle somme parziali, dovremo basarci su criteri per l'esistenza del limite, che non facciano intervenire la preliminare conoscenza del limite stesso. Essenzialmente, questi criteri si riducono a due soli: il teorema di Cauchy per le successioni e il teorema delle funzioni monotone. Esaminiamone le conseguenze per il caso delle serie.

1.3.1 Il teorema di Cauchy per le serie

Vediamo come si trascrive il Teorema di Cauchy nel caso in cui (s_n) è la successione delle somme parziali della serie

$$\sum_k x_k \,. \tag{1.3}$$

Sia, per fissare le idee, $m > n$. Allora,

$$|s_m - s_n| = \left| \sum_{k=n+1}^{m} x_k \right| .$$

Possiamo quindi enunciare il *Teorema di* *Cauchy* come segue:

Teorema 12 *La serie (1.3) converge se e solo se per ogni $\epsilon > 0$ esiste N_ϵ tale che per ogni coppia di indici n, m con*

$$m > n > N_\epsilon$$

vale

$$\left| \sum_{k=n+1}^{m} x_k \right| < \epsilon \,.$$

D'altra parte, notiamo che

$$\left| \sum_{k=n+1}^{m} x_k \right| \le \sum_{k=n+1}^{m} |x_k|$$

e quindi:

Corollario 13 *Se la serie*

$$\sum_k |x_k|$$

converge, anche la serie

$$\sum_k x_k$$

converge.

<u>Dim</u>. Infatti, se $\sum_k |x_k|$ converge, per ogni $\epsilon > 0$ esiste N_ϵ tale che per $m > n > N_\epsilon$ si ha

$$\left| \sum_{k=n+1}^{m} x_k \right| \leq \sum_{k=n+1}^{m} |x_k| < \epsilon \, .$$

E quindi anche la serie $\sum_k x_k$ converge, grazie al Teorema 12. ∎

Più avanti vedremo una diversa dimostrazione di questo corollario.

Si dice che la serie $\sum_k x_k$ *converge* *assolutamente* quando è convergente la serie $\sum_k |x_k|$. Il corollario precedente quindi può enunciarsi in questo modo:

Teorema 14 *Una serie assolutamente convergente è convergente.*

Questo risultato è molto importante perché la serie $\sum_k |x_k|$ è una serie a termini positivi. Criteri di convergenza facilmente usabili esistono appunto per il caso delle serie a termini positivi, come ora andiamo a vedere.

1.3.2 Monotonia e serie a termini di segno costante

Usando il teorema delle funzioni monotone, è facile vedere che

Teorema 15 *Sia (x_n) una successione a termini positivi. La serie degli x_n converge se e solo se esiste M tale che $s_n < M$ per ogni n.*

<u>Dim</u>. Ricordiamo il significato di $\sum_n x_n$: prima si costruisce la successione

$$s_k = \sum_{n=1}^{k} x_n$$

e poi si studia il limite $\lim_{k\to+\infty} s_k$.

La successione (s_k) è crescente perché, essendo $x_k \geq 0$ per ogni k,

$$s_{k+1} = \sum_{n=1}^{k+1} x_n = \left[\sum_{n=1}^{k} x_n \right] + x_{k+1} \geq \sum_{n=1}^{k} x_n = s_k \, .$$

Dunque la successione (s_k) ammette limite, finito o meno, per il *teorema delle funzioni monotone.* Il limite è finito se e solo se la successione (s_k) è superiormente limitata, ossia se e solo se esiste M tale che $s_k < M$ per ogni k. ∎

Il teorema facilmente si estende al caso di successioni a termini negativi oppure definitivamente positive o negative.

Inoltre:

Teorema 16 (Teorema del confronto) *siano $\sum x_n$ e $\sum y_n$ due serie a termini positivi, con $x_n \leq y_n$ per ogni n. Allora, se $\sum y_n$ converge, anche $\sum x_n$ converge; se $\sum x_n$ diverge lo stesso fa $\sum y_n$.*

Questo semplice risultato ha come conseguenza due importanti criteri di convergenza per le serie a termini positivi:

Teorema 17 (Criterio della radice) *Sia $x_n \geq 0$ per ogni n:*

- *Se esiste $q \in [0, 1)$ ed esiste N tale che*

$$\sqrt[n]{x_n} < q \qquad \forall n > N\,,$$

  allora la serie converge.

- *Se esiste $q > 1$ e se esiste una s.successione (x_{n_k}) tale che*

$$\sqrt[n_k]{x_{n_k}} > q$$

  allora la serie diverge.

<u>Dim</u>. Da $\sqrt[n]{x_n} < q < 1$ segue infatti $x_n < q^n$ e, se $0 \leq q < 1$, la convergenza della serie $\sum x_n$ segue dall'esempio 7 e dal Teorema del confronto.

Se per un $q > 1$ vale

$$\sqrt[n_k]{x_{n_k}} > q \qquad \text{ossia} \qquad x_{n_k} > q^{n_k}$$

allora[3]

$$\lim x_{n_k} = +\infty\,.$$

Di conseguenza il termine generale della serie non tende a zero, e quindi la serie non converge. ∎

Si ha inoltre:

[3]per provarlo si usi il teorema di confronto per i limiti.

Teorema 18 (Criterio del rapporto) *Se vale definitivamente*

$$\frac{x_{n+1}}{x_n} < q < 1 \tag{1.4}$$

allora $\sum x_n$ converge; *se* $\frac{x_{n+1}}{x_n} > q > 1$ *allora* $\sum x_n$ diverge.

<u>Dim</u>. Proviamo l'asserto nel caso in cui la (1.4) valga per ogni n.

Se $\frac{x_{n+1}}{x_n} < q < 1$ allora $x_2 < qx_1$, $x_3 < qx_2 < q^2x_1$ e, in generale, $x_n < q^{n-1}x_1$. Si sa che se $0 \le q < 1$ allora $\sum x_1 q^n = x_1 \sum q^n$ converge, si veda l'esempio 7. L'asserto segue quindi dal Teorema del confronto. In modo analogo si vede il secondo asserto. ∎

Ricordando i teoremi sui limiti, si può enunciare il corollario seguente:

Corollario 19 *Sia* $\sum x_n$ *una serie a termini positivi. Vale:*

- *se* $\lim_{n\to+\infty} \frac{x_{n+1}}{x_n} = q < 1$ *allora la serie converge;*
- *se* $\lim_{n\to+\infty} \sqrt[n]{x_n} = q < 1$ *allora la serie converge;*
- *se* $\lim_{n\to+\infty} \frac{x_{n+1}}{x_n} = q > 1$ *allora la serie diverge;*
- *se* $\lim_{n\to+\infty} \sqrt[n]{x_n} = q > 1$ *allora la serie diverge.*

Concludiamo con un esempio:

Esempio 20 Consideriamo la serie

$$\sum_{n=1}^{+\infty} \frac{1}{n}\,. \tag{1.5}$$

Mostriamo che questa serie è divergente.

Si noti che per ogni $x \ge -1$ vale

$$x \ge \log(1+x)\,.$$

Infatti, la funzione $\log(1+x)$ è concava e quindi ha grafico che sta sotto a ciascuna delle sue tangenti; e $y = x$ è la tangente nell'origine.

In particolare vale

$$\frac{1}{n} \ge \log\left(1+\frac{1}{n}\right)\,.$$

Abbiamo visto che la serie a termini positivi

$$\sum_{n=1}^{+\infty} \log\left(1+\frac{1}{n}\right)$$

diverge, si veda l'esempio 9. Dunque, per confronto, anche la serie (1.5) diverge[4]. ∎

La serie (1.5) si chiama *serie* *armonica*.

Si osservi che il carattere della serie armonica non può determinarsi usando il criterio del rapporto oppure quello della radice. Infatti, nel caso della serie armonica,

$$\lim \frac{x_{n+1}}{x_n} = \lim \frac{n}{n+1} = 1\,, \qquad \lim \sqrt[n]{x_n} = \lim \sqrt[n]{\frac{1}{n}} = 1\,.$$

All'esempio 24 vedremo una serie di termine generale x_n che è convergente e tale che anche per essa vale

$$\lim \frac{x_{n+1}}{x_n} = 1\,, \qquad \lim \sqrt[n]{x_n} = 1\,.$$

Combinando questi due esempi si ha:

niente può dedursi dai criteri della radice e del rapporto, se il numero q che compare in tali criteri è uguale ad 1.

Le serie a termini positivi hanno una notevole proprietà, che non è condivisa dalle generiche serie a termini di segno variabile: se si altera l'ordine di infiniti termini di una serie si trova una nuova serie, che generalmente ha un comportamento diverso da quello della serie di partenza. Invece:

Teorema 21 *Due serie a termini positivi, con gli stessi elementi in ordine diverso, hanno la medesima somma.*

La formula di Stirling

Per ragioni che vedremo, molto spesso il termine generale di una serie contiene dei fattoriali. I fattoriali hanno un "buon comportameno" rispetto al rapporto, nel senso che permettono facilmente di fare semplificazioni. Invece, il criterio della radice sembra difficile da usare in presenza dei fattoriali. In realtà non è così grazie alla *formula di Stirling*

$$n! \sim n^n e^{-n}\sqrt{2\pi n} \qquad \text{ossia} \qquad \lim \frac{n^n e^{-n}\sqrt{2\pi n}}{n!} = 1\,. \tag{1.6}$$

La dimostrazione si trova nei testi di Analisi Matematica 1.

[4] L'esempio 24 presenta una diversa dimostrazione di questo fatto.

1.3.3 Il test di MacLaurin

Consideriamo le somme parziali di una serie a termini positivi

$$\sum_{n=1}^{+\infty} a_n \,.$$

Esse sono

$$\begin{array}{lcl} s_1 = a_1 & = & a_1 \cdot 1 \\ s_2 = a_1 + a_2 & = & a_1 \cdot 1 + a_2 \cdot 1 \\ s_3 = a_1 + a_2 + a_3 & = & a_1 \cdot 1 + a_2 \cdot 1 + a_3 \cdot 1 \\ \vdots & & \end{array}$$

Queste espressioni si possono interpretare come somma di aree di rettangoli interpretando 1 come misura della base ed a_n come misura dell'altezza.

$$s_k = \sum_{n=1}^{k} a_n = \int_1^k a(x)\, \mathrm{d}x$$

ove $a(x)$ è la funzione costante a tratti

$$a(x) = a_n \quad \text{se} \quad 1 \le n \le x < (n+1)\,.$$

Dunque, la somma della serie è l'integrale improprio di $a(x)$:

$$\sum_{n=1}^{+\infty} a_n = \lim_{k\to+\infty} s_k = \int_1^{+\infty} a(x)\, \mathrm{d}x\,.$$

Pensiamo ora ai rettangoli messi come in figura 1.1, a sinistra, e supponiamo di poter trovare due funzioni, $f(x)$ e $g(x)$, che prendono valori maggiori o uguali a zero e tali che inoltre valga

$$x \in [n, n+1) \implies g(x) \le a_n \le f(x)\,.$$

Si veda la figura 1.1, a destra. In tal caso si ha

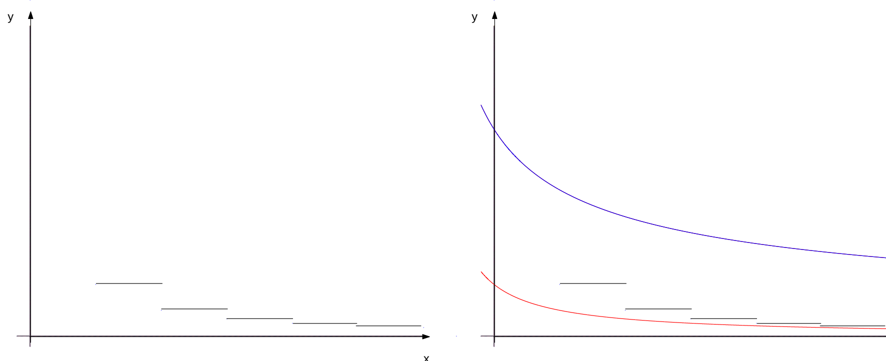
$$\int_1^k g(x)\, \mathrm{d}x \le s_k \le \int_1^k f(x)\, \mathrm{d}x\,.$$

La serie è a termini positivi e quindi regolare; le funzioni sono non negative e quindi ammettono integrale improprio finito o meno. Dunque, dal teorema di confronto per i limiti, si ha

$$\int_0^{+\infty} g(x)\, \mathrm{d}x \le \sum_{n=1}^{+\infty} a_n \le \int_0^{+\infty} f(x)\, \mathrm{d}x\,.$$

Ricapitolando,

Figura 1.1: Il test di MacLaurin



- Se $\int_1^{+\infty} g(x)\,\mathrm{d}x = +\infty$ allora la serie diverge;
- Se $\int_1^{+\infty} f(x)\,\mathrm{d}x < +\infty$ allora la serie converge. In questo caso si trovano anche stime, da sotto e da sopra, per la somma della serie.

Il caso tipico in cui quest'argomento si applica facilmente è il caso in cui esiste una funzione $g(x)$ definita su $[0,+\infty)$, decrescente e inoltre

$$a_n = g(n)\,.$$

In questo caso,

$$x \in [n, n+1) \implies g(x) \le a_n = g(n) \le g(x-1)\,. \tag{1.7}$$

Definiamo, per $x \ge 0$,

$$f(x) = g(x-1)$$

e notiamo che la (1.7) si scrive

$$g(x) \le a_n = g(n) \le f(x) \qquad x \in [n, n+1)\,.$$

Inoltre,

$$\int_1^{+\infty} f(x)\,\mathrm{d}x = \int_1^{+\infty} g(x-1)\,\mathrm{d}x = \int_0^1 g(x)\,\mathrm{d}x + \int_1^{+\infty} g(x)\,\mathrm{d}x\,,$$

$$\int_1^{+\infty} g(x)\,\mathrm{d}x < +\infty \iff \int_1^{+\infty} f(x)\,\mathrm{d}x < +\infty\,.$$

Ossia, nel caso descritto, i due integrali impropri hanno il medesimo comportamento e questo comportamento è ereditato dalla serie. Possiamo quindi enunciare:

Teorema 22 (*Test di MacLaurin*) *Sia* $g(x)$ *una funzione* non negativa e decrescente *definita su* $[0,+\infty)$*. Si consideri la serie*

$$\sum_{n=1}^{+\infty} g(n)\,.$$

Le sue somme parziali verificano

$$\int_1^k g(x)\,dx \le s_k = \sum_{n=1}^{k} a_n \le \int_0^1 g(x)\,dx + \int_1^k g(s)\,ds\,. \tag{1.8}$$

In particolare, la serie a termini positivi

$$\sum_{n=1}^{+\infty} g(n)$$

converge se e solo se

$$\int_0^{+\infty} g(x)\,dx < +\infty\,.$$

L'interesse di questo teorema sta nel fatto che talvolta l'integrale di $g(x)$ può esplicitamente calcolarsi mediante il calcolo delle primitive; e comunque esistono test efficienti per lo studio della convergenza o divergenza degli integrali impropri.

Esempio 23 Si sa già che la serie armonica

$$\sum_{n=1}^{+\infty} \frac{1}{n}$$

diverge. La serie

$$\sum_{n=2}^{+\infty} \frac{1}{n\log^2 n}$$

converge, come si vede dal criterio di MacLaurin. Infatti, la funzione

$$f(x) = \frac{1}{x\log^2 x}$$

ha integrale improprio convergente:

$$\lim_{T\to+\infty} \int_2^T \frac{1}{x\log^2 x}\,\mathrm{d}x = \lim_{T\to+\infty}\left[\frac{1}{\log 2} - \frac{1}{\log T}\right] = \frac{1}{\log 2}\,.$$

Procedendo in modo analogo[5] si provi invece che

$$\sum_{n=2}^{+\infty} \frac{1}{n \log n} = +\infty \, . \quad \blacksquare$$

Esempio 24 Si calcola immediatamente

$$\int_0^{+\infty} \frac{1}{(x+1)^\gamma} \, \mathrm{d}x = \begin{cases} < +\infty & \text{se } \gamma > 1 \\ = +\infty & \text{se } \gamma \geq 1. \end{cases}$$

Dunque,

$$\sum_{n=1}^{+\infty} \frac{1}{n^\gamma}$$

converge per $\gamma > 1$, diverge altrimenti. ∎

Possiamo combinare l'esempio 24 col criterio di confronto, ottenendo:

Corollario 25 *Consideriamo la serie* $\sum_{n=1}^{+\infty} a_n$. *Vale:*

- *se esitono* $M > 0$ *e* $\gamma > 1$ *tali che*

  $$0 \leq a_n \leq M \frac{1}{n^\gamma} \, ,$$

  allora la serie converge.

- *se esistono* $m > 0$ *e* $\gamma \leq 1$ *tali che*

  $$a_n \geq m \frac{1}{n^\gamma}$$

  allora la serie diverge.

In particolare, possiamo enunciare:

> Se $a_n \geq 0$ e se esiste $\gamma > 1$ tale che
>
> $$a_n = \mathrm{o}\left(\frac{1}{n^\gamma}\right) \, , \qquad (1.9)$$
>
> allora la serie $\sum a_n$ converge.

Per ora, stiamo lavorando con serie a termini positivi, ma non abbiamo scritto esplicitamente questa condizione perché vedremo, al Corollario 27, che il test precedente vale per ogni serie.

[5] si usi

$$\frac{\mathrm{d}}{\mathrm{d}x} \log[\log x] = \frac{1}{x \log x} \, .$$

1.3.4 Serie a termini di segno qualsiasi

Sulle serie a termini di segno qualsiasi, limitiamoci ad osservare due proprietà.

Si è già detto che se la serie $\sum |x_n|$ converge, si dice che la serie $\sum x_n$ *converge* *assolutamente*. Ricordiamo, dal teorema 15:

Teorema 26 *Una serie assolutamente convergente è convergente.*

Ricordiamo ora che $f = \mathrm{o}(g)$ quando f/g è un infinitesimo, e ciò accade se e solo se $|f|/|g|$ è un infinitesimo. Quindi:

Corollario 27 *Se esite $\gamma > 1$ tale che $a_n = \mathrm{o}\left(\frac{1}{n}\right)$, allora la serie $\sum a_n$ converge assolutamente, ed è quindi convergente.*

Infine, si dice che una serie è *a segni* *alterni* se ha forma

$$\sum (-1)^n x_n \qquad \text{con } x_n > 0; \tag{1.10}$$

ossia se gli addendi si susseguono cambiando segno ad ogni passo. Esiste, per le serie a segni alterni, una notevole condizione sufficiente di convergenza, e anche una stima per la somma della serie:

Teorema 28 (*Criterio di Leibniz*) *Se valgono ambedue le condizioni*

- $\lim_{n\to+\infty} x_n = 0$
- *la succesione* $\{x_n\}$ è decrescente, *ossia* $x_n \geq x_{n+1} \geq 0$ *per ogni* n

allora la serie a segni alterni (1.10) converge; inoltre, detta s la somma della serie, per ogni n vale:

- *la differenza*

  $$s - \sum_{n=1}^{k} (-1)^n x_n$$

  ha segno opposto ad x_k; ossia, l'approssimazione è per eccesso se l'ultimo termine sommato è positivo; per difetto se è negativo.

- *Vale la stima*

  $$\left| s - \sum_{n=1}^{k} x_n \right| \leq |x_{k+1}| \,.$$

La dimostrazione della convergenza è nell'Appendice 1.8.2.

Esempio 29 Consideriamo la serie

$$\sum_{n=1}^{+\infty} \frac{(-1)^n}{n} \, .$$

Questa serie si chiama *serie di Mengoli*. Il criterio di Leibniz mostra che questa serie converge. La somma della serie è nota:

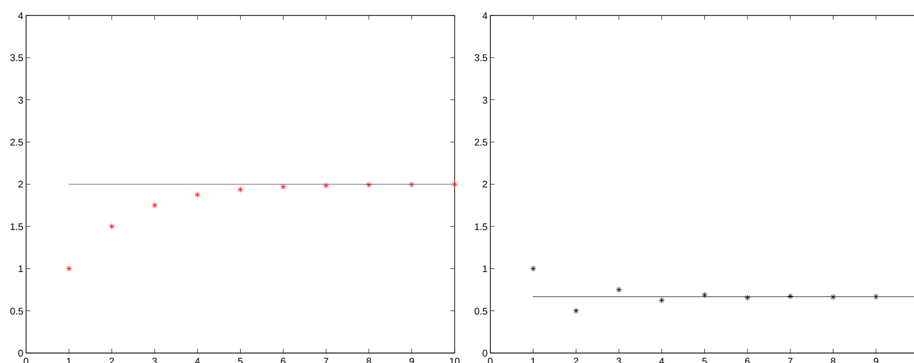
$$\sum_{n=1}^{+\infty} \frac{(-1)^n}{n} = \log 2 \, .$$

Il Teorema 28 dà anche una stima dell'errore che si commette sommando N termini: l'errore è minore di $1/(n+1)$. ∎

1.4 Alcuni esempi numerici

Le figure 1.2, 1.3 e 1.4 mostrano alcuni esempi numerici di somme parziali di serie convergenti. Le serie sono specificate nelle intestazioni delle figure.

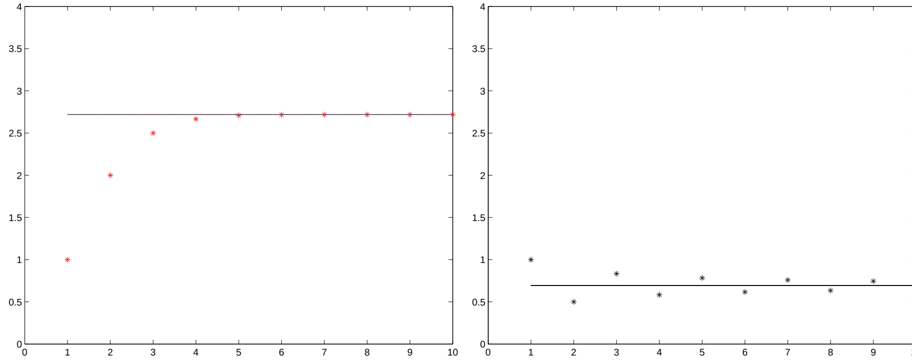
Figura 1.2: $\sum_{n=0}^{+\infty}(1/2)^n = 2$ a sinistra, $\sum_{n=0}^{+\infty}(-1/2)^n = 2/3$ a destra



1.5 Convergenza condizionata ed incondizionata

Il concetto di serie generalizza quello di somma finita. In una somma finita il risultato non dipende dall'ordine degli addendi (proprietà commutativa dell'addizione). Nel caso delle serie, l'asserto analogo vale se si scambiano tra di loro

Figura 1.3: $\sum_{n=0}^{+\infty} 1/n! = e$ a sinistra, $\sum_{n=0}^{+\infty}(-1)^n/n = \log 2$ a destra



un numero finito di termini. E' FALSO se si scambia il posto di infiniti termini. Mostriamo un esempio:

Esempio 30 Consideriamo la serie di Mengoli $\sum_{n=1}^{+\infty}(-1)^n/n$. Si sa che questa serie converge. Mostriamo che è possibile scambiare il posto di infiniti termini, in modo da ottenere una serie divergente a $+\infty$. Ricordiamo per questo che

$$\sum_{n=1}^{+\infty} \frac{1}{n}$$

diverge e quindi anche

$$\sum_{n=1}^{+\infty} \frac{1}{2n}$$

diverge. Conviene vedere una dimostrazione di questo fatto, diversa da quella già vista: consideriamo

$$\sum_{n=k}^{5k} \frac{1}{2n} = \frac{1}{2} \sum_{n=k}^{5k} \frac{1}{n} \geq \frac{1}{2}[4k \frac{1}{5k}] = \frac{2}{5}$$

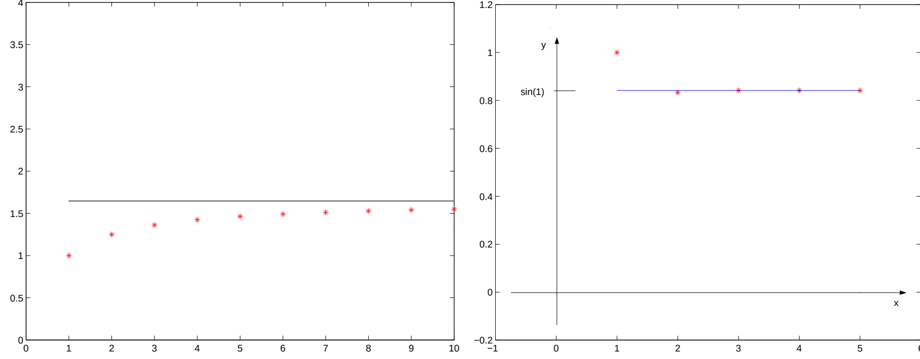
(si è usato il fatto che si sommano $4k$ termini, ciascuno dei quali è maggiore di $1/5k$).

Ciò contrasta col criterio di convergenza di Cauchy, e mostra che la serie diverge.

Dato che il carattere di una serie non dipende dai primi elementi, anche ciascuna delle serie

$$\sum_{n=R}^{+\infty} \frac{1}{2n} \tag{1.11}$$

Figura 1.4: $\sum_{n=1}^{+\infty} 1/n^2 = \pi^2/6$ a sinistra e $\sum_{n=0}^{+\infty}(-1)^n/[(2n+1)!\,] = \sin(1)$ a destra



è divergente.

Ora consideriamo la serie di Mengoli, per semplicità cambiata di segno. Essa converge a $-\log 2$. Vogliamo riordinarne gli elementi in modo da trovare una serie divergente a $+\infty$. Per questo sommiamo prima i termini di indice pari, fino ad un certo indice $\tilde{k}_1$ tale che

$$\sum_{n=1}^{\tilde{k}_1} \frac{1}{2n} > 5\,.$$

Questa è la somma parziale $s_{\tilde{k}_1}$ della serie riordinata. Sottraiamo quindi il primo termine di indice dispari, ossia 1 ottenendo una somma parziale s_{k_1} tale che

$$s_{k_1} > 4\,.$$

Consideriamo ora la serie (1.11) con $R = \tilde{k}_1 + 1$. Come si è detto, questa serie diverge. Dunque, possiamo sommare ulteriori termini di indice pari alla somma parziale s_{k_1} già ottenuta, fino a trovare una somma parziale maggiore di 9; sottraiamo quindi il primo dei termini di ordine dispari non ancora usati (che è $1/3$, minore di 1). Si trova una nuova somma parziale, diciamo s_{k_2}, maggiore di 8:

$$s_{k_2} > 8 = 2^3\,.$$

Continuiamo a sommare termini di indice pari (e quindi positivi) fino ad avere una somma parziale maggiore di 2^4+1 e quindi sottraiamo il primo termine di indice dispari non usato (che è certamente minore di 1, infatti è $1/5$). Si trova una somma parziale s_{k_3} tale che

$$s_{k_3} > 2^4\,.$$

Procedendo in questo modo si trova un riordinamento che conduce ad una serie divergente a $+\infty$. ∎

Si potrebbe mostrare che per ogni scelta di l è possibile riordinare la serie di Mengoli in modo tale da trovare una serie convergente ad l, inclusi $l = +\infty$ ed $l = -\infty$, o anche in modo da trovare una serie oscillante.

Diciamo che una serie *converge incondizionatamente* quando una serie converge ad l e inoltre quando qualunque serie ottenuta riordinandone gli elementi converge al medesimo numero l.

La convergenza incondizionata si caratterizza come segue:

Teorema 31 (***Teorema di Dirichlet***) *Una serie converge incondizionatamente se e solo se converge assolutamente.*

Se ciò non accade è possibile riordinare gli elementi della serie in modo da cambiare il carattere della serie, e anche in modo da ottenere una serie convergente ad un qualsiasi numero assegnato, o divergente a $+\infty$ oppure a $-\infty$.

In particolare:

Corollario 32 *Ogni serie a termini di segno costante converge incondizionatamente.*

1.5.1 Serie dipendenti da un parametro e serie di funzioni

Torniamo a considerare la serie geometrica,

$$\sum_{n=0}^{+\infty} q^n \,.$$

Questa serie dipende dal parametro q e, come si è visto, converge se $|q| < 1$, diverge se $q \geq 1$ ed oscilla se $q \leq -1$.

Facendo variare il parametro q, ciascuno degli addendi viene ad essere una funzione di q,

$$f_n(q) = q^n \,.$$

Dunque, la serie geometrica può intendersi come serie di funzioni. In generale, data la successione $(\, f_n(x)\,)$ i cui elementi sono funzioni (tutte con lo stesso dominio) si chiama *serie di funzioni* la serie

$$\sum_{n=0}^{+\infty} f_n(x) \,.$$

La somma della serie si calcola punto per punto; ossia, per ogni fissato valore di x si calcola la somma della serie di numeri $\sum_{n=0}^{+\infty} f_n(x)$.

Il dominio comune alle funzioni $f(q) = q^n$ è $\mathbf{R}$, ma abbiamo notato che la serie geometrica converge (ad $1/(1-q)$) soltanto per $|q| < 1$. Dunque, in generale, il dominio su cui è definita la somma di una serie di funzioni è più piccolo del dominio comune delle funzioni.

Le serie di funzioni si studieranno al Capitolo 2. Va tenuta presente la loro definizione per capire alcune sottigliezze del paragrafo 1.6.

1.6 Operazioni algebriche e serie

Il concetto di "serie" estende quello di "somma finita". Le somme finite godono di utili proprietà, come per esempio la proprietà distributiva del prodotto sulla somma, la proprietà associativa e "dissociativa". Ci possiamo chiedere se le analoghe proprietà valgono per le serie. Per questo dobbiamo tener conto di due problemi:

- la "somma di somme finite" non dipende dall'ordine degli addendi:

  $$(a_1 + a_2) + (b_1 + b_2) = a_1 + a_2 + b_1 + b_2 = a_1 + b_1 + a_2 + b_2 \,.$$

  Si sa già che la somma della serie varia cambiando l'ordine dei suoi termini. Quindi dovremo aspettarci che una proprietà analoga non valga necessariamente per le serie.

- la somma di serie è definita tramite il concetto di limite; le relazioni tra limiti ed operazioni sono dissimmetriche. Di ciò dobbiamo tener conto per enunciare i risultati relativi alle serie.

Queste osservazioni suggeriscono di elencare prima le relazioni tra limiti ed operazioni sia nella versione "giusta" che nella versione "sbagliata":

GIUSTA	**SBAGLIATA**
1 La somma dei limiti è uguale al limite della somma	**1** Il limite della somma è uguale alla somma dei limiti
2 il prodotto dei limiti è uguale al limite del prodotto	**2** il limite del prodotto è uguale al prodotto dei limiti
3a Sia $\{\gamma_n\}$ limitata. Se $a_n \to 0$ anche $\gamma_n a_n \to 0$.	**3a** Sia $\{\gamma_n\}$ limitata. Se $\{\gamma_n a_n\}$ converge, anche $\{a_n\}$ converge.
3b Sia $\{\gamma_n\}$ limitata. Se $a_n \to 0$ anche $\gamma_n a_n \to 0$.	**3b** Sia $\{\gamma_n\}$ limitata. Se $\{a_n\}$ converge, anche $\{\gamma_n a_n\}$ converge.

Ricordiamo brevemente perché le affermazioni a destra sono sbagliate: per esempio nel caso **1**, il limite della somma di due funzioni può esistere, senza che le due funzioni individualmente abbiano limite, come è il caso del limite, per $x \to +\infty$, quando le due funzioni sono

$$f(x) = x^3 + \sin x\,, \qquad g(x) = \frac{x+1}{1-x}\sin x\,.$$

In questo caso,

$$f(x) + g(x) = \frac{2\sin x}{1-x}\,, \qquad \lim_{x\to+\infty}[f(x)+g(x)] = 0\,.$$

Questo limite non è uguale a

$$\left[\lim_{x\to+\infty} f(x)\right] + \left[\lim_{x\to+\infty} g(x)\right]$$

per la semplice ragione che i due limiti non esistono.

Il caso **2**, del prodotto è analogo. Vale la pena però di vedere un caso banale che però può inndurre in errore nel caso delle serie: NON è vero che la formula seguente vale per OGNI numero reale γ:

$$\lim\,[\gamma f(x)] = \gamma\,[\lim f(x)]\ .$$

Questa formula vale solo se $\gamma \neq 0$ oppure se $\lim f(x)$ esiste finito, come mostrano gli esempi delle funzioni

$$f(x) = x^3\,, \qquad g(x) = \sin x\,.$$

In ambedue i casi, se $\gamma = 0$,

$$\lim_{x\to+\infty}[\gamma f(x)] = 0\,, \qquad \lim_{x\to+\infty}[\gamma g(x)] = 0$$

mentre le espressioni

$$\gamma\lim_{x\to+\infty} f(x)\,, \qquad \gamma\lim_{x\to+\infty} g(x)$$

non hanno senso.

Vediamo ora il caso **3a**. Chiaramente, da $\gamma_n a_n \to 0$ non si può dedurre la convergenza a zero di $\{a_n\}$: si consideri il caso $\gamma_n \to 0$ ed $a_n \equiv 1$.

Il caso **3b**: si consideri l'esempio della successione $\{a_n\}$ con $a_n = (n-1)/(n+1)$, convergente ad 1, mentre $\gamma_n = (-1)^n$. La successione $\{\gamma_n a_n\}$ è oscillante.

Richiamato ció, definiamo:

Somma di serie

$$\left[\sum_{n=1}^{+\infty} a_n\right] + \left[\sum_{n=1}^{+\infty} b_n\right] = \sum_{n=1}^{+\infty}(a_n + b_n)\,.$$

Prodotto di una serie per un numero

$$\alpha \cdot \left[\sum_{n=1}^{+\infty} a_n\right] = \sum_{n=1}^{+\infty}(\alpha a_n)\,.$$

In queste definizioni, l'ordine degli addendi è quello indicato.

Supponiamo

$$\sum_{n=1}^{+\infty} a_n = l\,, \qquad \sum_{n=1}^{+\infty} b_n = m\,.$$

Allora:

1) somma di serie. Vale

$$\sum_{n=1}^{+\infty}(a_n + b_n) = l + m$$

purché uno almeno dei due limiti sia un numero oppure sia l che m siano ambedue $+\infty$ oppure $-\infty$.

2) prodotto di un numero per una serie. La definizione ha senso (e l'uguaglianza vale) solo se $\alpha \neq 0$ oppure se la serie converge.

La definizione di prodotto di serie è più complessa e si vedrà nel paragrafo 1.7.

Si potrebbe anche provare:

Teorema 33 *Sia* $\sum a_n = l \in \mathbf{R}$ *e sia* $a_n > 0$ *per ogni* n. *Sia* $\{\gamma_n\}$ *una successione* limitata. *Allora, la serie* $\sum \gamma_n a_n$ *converge.*

Gli esempi seguenti mostrano i problemi che si possono incontrare usando le operazioni sulle serie senza le dovute cautele:

Esempio 34 Consideriamo la serie seguente:

$$\sum c_n\,, \qquad c_n = 0\,.$$

Ovviamente la somma della serie è 0. Scrivendo

$$c_n = a_n + b_n\,, \qquad a_n = (-1)^n\,, \quad b_n = -(-1)^n$$

si potrebbe essere tentati di usare una specie di “regola dello scomponendo” e scrivere

$$\sum c_n = \sum a_n + \sum b_n \,.$$

Ovviamente questo non ha senso, perché le due serie a destra non convergono; e quindi non definiscono numeri che si possano sommare.

Un esempio analogo, un po’ più riposto, è quello delle serie

$$\sum a_n \,, \quad \sum b_n \,, \qquad a_n = \frac{1}{n+1} \,, \quad b_n = \frac{n\left(\sqrt{n}-n\right)}{n^3+8} \,.$$

Le due serie non convergono mentre la serie

$$\sum [a_n + b_n]$$

converge. ∎

La regula del prodotto sembra “più innocua” nel senso che sembra più difficile sbagliare. In realtà anche questa regola è fonte di errori, come mostra l’esempio seguente:

Esempio 35 Sia x un parametro reale. Studiare la convergenza della serie

$$\sum_{n=1}^{+\infty} \frac{x}{n^{1/2}(1+nx^2)} \,. \tag{1.12}$$

Si lascia per esercizio di provare che la serie converge per ogni x. Si noti che essa certamente converge per $x = 0$ perché in tal caso tutti i termini della serie sono nulli. Però, sembra del tutto naturale mettere in evidenza x portandolo fuori dal segno di serie, scrivendo

$$x \left[\sum_{n=1}^{+\infty} \frac{1}{n^{1/2}(1+nx^2)} \right]$$

e magari studiando la convergenza della serie “piú semplice”

$$\sum_{n=1}^{+\infty} \frac{1}{n^{1/2}(1+nx^2)} \,.$$

Si dimentica in questo modo che il parametro x può essere nullo. Se $x \neq 0$ questa serie converge e il procedimento seguito, di mettere in evidenza x, è corretto. Se però $x = 0$, l’ultima serie scritta non converge, mentre la (1.12) ovviamente converge.

L’errore è consistito nel “mettere in evidenza” il fattore 0 dai termini della serie, errore favorito dal fatto che il fattore è stato indicato col generico simbolo x. ∎

Ci sono anche altri errori nei quali si può cadere trattando le operazioni sulle serie senza la dovuta attenzione:

Esempio 36 Un modo veloce di “calcolare” la somma della serie geometrica

$$S = \sum_{n=0}^{+\infty} q^n$$

è il seguente:

$$S = 1 + q\left[\sum_{n=0}^{+\infty} q^n\right] = 1 + qS$$

Dunque,

$$(1-q)S = 1 \qquad \text{ossia} \qquad S = \frac{1}{1-q}\,.$$

Questo risultato, apparentemente giusto, è in realtà sbagliato. Infatti da nessuna parte si è usata la condizione $|q| < 1$ e quindi sembrerebbe da questo calcolo che la serie geometrica converga per ogni valore di q, cosa notoriamente falsa.

Si lascia per esercizio di trovare l'errore in questo ragionamento. ∎

Infine, vediamo una diversa dimostrazione del Teorema 26 basata sulle proprietà illustrate in questo paragrafo:

La serie $\sum |x_n|$ è una serie a termini positivi. Se essa converge, dal teorema 15 convergono anche le due serie $\sum y_n$ e $\sum z_n$, con

$$y_n = \begin{cases} x_n & \text{se} \quad x_n > 0 \\ 0 & \quad \text{altrimenti} \end{cases}, \qquad z_n = \begin{cases} -x_n & \text{se} \quad x_n < 0 \\ 0 & \quad \text{altrimenti.} \end{cases}$$

Quindi, per il teorema 16, converge anche $\sum(y_n - z_n)$ che è $\sum x_n$.

1.7 Prodotto alla Cauchy

Il prodotto di serie può definirsi in vari modi. Quello più utile è il *prodotto alla Cauchy*. L'espressione del prodotto alla Cauchy può sembrare macchinosa, ma se ne capisce la ragione se si considera l'esempio di un prodotto di polinomi. Consideriamo il caso del prodotto di due polinomi di grado 4. Il prodotto

$$(a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4)(b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4)$$

è la somma di tutti i possibili addendi $a_jb_kx^{k+j}$ con $0 \le j \le 4$, $0 \le k \le 4$. Raccogliendo i coefficienti degli addendi del medesimo grado il prodotto si scrive come

$$\begin{aligned}
&a_0b_0\\
&+(a_1b_0 + b_0a_1)x\\
&+(a_2b_0 + a_1b_1 + a_0b_2)x^2\\
&+(a_3b_0 + a_2b_1 + a_1b_2 + a_0b_3)x^3\\
&+(a_4b_0 + a_3b_1 + a_2b_2 + a_1b_3 + a_0b_4)x^4\,.
\end{aligned}$$

Posto $x = 1$, il prodotto

$$(a_0 + a_1 + a_2 + a_3 + a_4)(b_0 + b_1 + b_2 + b_3 + b_4)$$

si trova scritto come segue:

$$a_0b_0$$

$$
\begin{aligned}
&+(a_1b_0+b_0a_1)\\
&+(a_2b_0+a_1b_1+a_0b_2)\\
&+(a_3b_0+a_2b_1+a_1b_2+a_0b_3)\\
&+(a_4b_0+a_3b_1+a_2b_2+a_1b_3+a_0b_4)\,.
\end{aligned}
$$

Ciò suggerisce le due definizioni seguenti:

- Si chiama *convoluzione* delle due successioni (a_n) e (b_n) la successione (c_n) con

$$
c_n=\sum_{r=0}^{n} a_{n-r}b_r = a_nb_0+a_{n-1}b_1+a_{n-2}b_2+\cdots a_2b_{n-2}+a_1b_{n-1}+a_0b_n\,.
$$

- Si chiama *prodotto alla Cauchy* delle due serie

$$
\sum_{n=0}^{+\infty} a_n\,,\qquad \sum_{n=0}^{+\infty} b_n \tag{1.13}
$$

  la serie

$$
\sum_{n=0}^{+\infty} c_n\,,\qquad c_n=\sum_{r=0}^{n} a_{n-r}b_r\,.
$$

Osservazione 37 Se l'indice di una, o ambedue, le serie non parte da zero, la formula del prodotto alla Cauchy si intende scritta aggiungendo un numero finito di termini tutti nulli, in modo da far partire gli indici da 0. ∎

La convergenza di ambedue le serie (1.13) non implica la convergenza del loro prodotto alla Cauchy. Vale invece:

Teorema 38 *Le due serie (1.13) convergano, ed abbiano somma rispettivamente α e β. Allora:*

- *se le due serie convergono ambedue assolutamente, anche il loro prodotto alla Cauchy converge assolutamente ad $\alpha\beta$.*
- *se una delle due serie converge e l'altra converge assolutamente, il prodotto alla Cauchy converge ad $\alpha\beta$, in generale non assolutamente.*

Concludiamo con un esempio che mostra due serie convergenti (non assolutamente), il cui prodotto alla Cauchy non converge.

Esempio 39 Consideriamo la serie (convergente per il criterio di Leibniz)

$$
\sum_{n=1}^{+\infty}\frac{(-1)^n}{\sqrt{n}}
$$

e calcoliamone il prodotto alla Cauchy con se stessa. Notiamo che l'indice di questa serie parte da 1 invece che da zero e quindi la formula del prodotto alla Cauchy va lievemente modificata come detto nell'sservazione 37:

$$
\sum_{n=1}^{+\infty}\left[\sum_{r=1}^{n-1}\frac{(-1)^{n-r}}{\sqrt{n-r}}\frac{(-1)^r}{\sqrt{r}}\right]=\sum_{n=1}^{+\infty}(-1)^n\left[\sum_{r=1}^{n-1}\frac{1}{\sqrt{nr-r^2}}\right]\,.
$$

Si vede facilmente che la funzione

$$f(x)=\frac{1}{\sqrt{nx-x^2}}$$

è crescente per $1<x<n/2$ e decrescente per $n/2<x<n-1$ e quindi ha minimo per $x=1$ e per $x=n-1$. Il minimo vale $1/\sqrt{n-1}$. Dunque

$$\sum_{r=1}^{n-1}\frac{1}{\sqrt{nr-r^2}}\geq\frac{n-2}{\sqrt{n-1}}\ :$$

il termine generale del prodotto alla Cauchy non converge a zero e quindi la serie ottenuta come prodotto alla Cauchy non converge. ∎

1.8 Appendici

1.8.1 Appendice: ancora sul test di MacLaurin

Il test di MacLaurin può ulteriormente precisarsi notando che la (1.8), ossia

$$\int_1^k g(x)\,\mathrm{d}x\leq s_k=\sum_{n=1}^k a_n\leq\int_0^1 g(x)\,\mathrm{d}x++\int_1^k g(x)\,\mathrm{d}x\,,$$

si può anche scrivere come

$$0\leq\left[\sum_{n=1}^k a_n\right]-\int_1^k g(x)\,\mathrm{d}x\leq\int_0^1 g(x)\,\mathrm{d}x$$

e che, al crescere di k, la successione

$$k\longrightarrow\left\{\left[\sum_{n=1}^k a_n\right]-\int_1^k g(x)\,\mathrm{d}x\right\}$$

decresce. e quindi ammette limite compreso tra 0 ed $\int_0^1 g(x)\,\mathrm{d}x$. Infatti, si ha:

$$\left\{\left[\sum_{n=1}^k a_n\right]-\int_1^k g(x)\,\mathrm{d}x\right\}-\left\{\left[\sum_{n=1}^{k+1} a_n\right]-\int_1^{k+1} g(x)\,\mathrm{d}x\right\}$$
$$=-a_{k+1}+\int_k^{k+1} g(x)\,\mathrm{d}x=\int_k^{k+1}\left[g(x)-g(k+1)\right]\,\mathrm{d}x\geq 0\,.$$

Ossia, al crescere di k, i valori delle somme parziali e dell'integrale "si avvicinano" anche se l'integrale improprio (e quindi anche la serie) diverge. Quest'osservazione può usarsi per ottenere "stime asintotiche" delle somme parziali della serie per grandi valori di k. Mostriamo questo considerando l'esempio della serie armonica.

Sia

$$g(x)=\begin{cases}1 & \text{se}\quad 0\leq x\leq 1\\ 1/x & \text{se}\quad 0\leq x>1\,.\end{cases}$$

La serie armonica è

$$\sum_{n=1}^{+\infty} g(n)$$

e quindi diverge, dal criterio di MacLaurin. E' questo un modo di vedere che la serie armonica diverge, diverso da quello visto all'esempio 1.5. Però, il criterio di MacLaurin dà un'informazione in più:

$$0 \le \sum_{n=1}^{k} \frac{1}{n} - \int_1^k \frac{1}{x}\,\mathrm{d}x \le \int_0^1 g(x)\,\mathrm{d}x\,,$$

ossia

$$0 \le \sum_{n=1}^{k} \frac{1}{n} - \log k \le 1\,.$$

E inoltre,

$$\gamma = \lim_k \left[\sum_{n=1}^{k} \frac{1}{n} - \log k \right]$$

esiste, $\gamma \in (0,1)$. Il numero γ così definito si chiama *costante d'Eulero*.

1.8.2 La dimostrazione del Teorema di Leibniz

La dimostrazione di questo teorema, e di teoremi più generali di Abel e di Dirichlet che ora vedremo, è interessante perché permette di introdurre il concetto di sommazione per parti, da confrontare con quello di integrazione per parti.

Nonostante che il carattere di una serie non dipenda dai primi addendi, in quest'appendice è bene specificare con cura i valori degli indici. Quindi scriveremo per esempio $\{b_n\}_{n\ge1}$ per intendere che il valore del primo indice della successione è 1. Inoltre, se $\{b_n\}_{n\ge1}$ indica una successione, con $\{B_n\}_{n\ge1}$ intendiamo la successione delle sue somme parziali:

$$B_n = \sum_{k=1}^{n} b_k\,.$$

Sia $\{b_n\}_{n\ge1}$ una successione. Con $\{\Delta b_n\}_{n\ge1}$ intendiamo la successione $\{(b_{n+1}-b_n)\}_{n\ge1}$.

Sia $1 \le r < k$. La formula di *sommazione per parti* è:

$$\sum_{n=r}^{k} (\Delta b_n) c_n = [b_{k+1}c_k - b_r c_r] - \sum_{n=r}^{k} b_{n+1}\Delta c_n\,.$$

Questa formula si dimostra facilmente scrivendo la somma per esteso:

$$\sum_{n=r}^{k} (\Delta b_n) c_n =$$

$$(b_{r+1}-b_r)c_r + (b_{r+2}-b_{r+1})c_{r+1} + (b_{r+3}-b_{r+2})c_{r+2} + \cdots + (b_k - b_{k-1})c_{k-1} + (b_{k+1}-b_k)c_k$$

$$= -b_r c_r + [-b_{r+1}(c_{r+1}-c_r) - b_{r+2}(c_{r+2}-c_{r+1}) - \cdots - b_k(c_k - c_{k-1})] + b_{k+1}c_k$$

$$= b_{k+1}c_k - b_r c_r - \sum_{n=r}^{k-1} b_{n+1}\Delta c_n = -b_{k+1}\left[c_{k+1} - c_k\right] + b_{k+1}c_{k+1} - b_r c_r - \sum_{n=r}^{k-1} b_{n+1}\Delta c_n$$

$$= b_{k+1}c_{k+1} - b_r c_r - \sum_{n=r}^{k} b_{n+1}\Delta c_n\,.$$

Ci serve inoltre un risultato preliminare:

Lemma 40 *Sia* $\sum_{n=1}^{+\infty} b_n$ *una serie tale che la successione* $\{B_n\}$ *delle somme parziali sia* limitata *e sia* $\{c_n\}$ *una successione* positiva decrescente. *Allora, la serie*

$$\sum_{n=1}^{+\infty} B_n(\Delta c_n)$$

converge assolutamente.

Dim. Si noti che la successione $\{c_n\}_{n\geq 1}$ è convergente, $\lim c_n = l$, per il teorema della funzione monotona, e quindi limitata. Anzi,

$$0 \leq c_n \leq c_1\,.$$

Inoltre, $c_{n+1} - c_n < 0$.

L'asserto del lemma segue perché ora proviamo che $\sum_{n=1}^{+\infty} |B_n\Delta c_n| < +\infty$. Infatti,

$$\sum_{n=1}^{k} |B_n\Delta c_n| = \sum_{n=1}^{k} |B_n|\,(c_n - c_{n+1}) \leq M\sum_{n=1}^{k}(c_n - c_{n+1}) = M\,(c_1 - c_{k+1}) \leq Mc_1\,. \quad \blacksquare$$

Osservazione 41 Si noti che il Lemma 40 non richiede la convergenza della serie $\sum_{n=1}^{+\infty} b_n$. ∎

Consideriamo ora la serie $\sum_{n=1}^{+\infty} b_n c_n$ e notiamo che

$$b_n = B_n - B_{n-1} = \Delta B_{n-1}\,.$$

Introducendo i numeri d_n definiti da

$$d_{n-1} = c_n \qquad \text{ossia} \qquad d_n = c_{n+1}$$

e usando la regola di sommazione per parti con primo indice 2, le sue somme parziali si rappresentano come segue:

$$\sum_{n=1}^{k} b_n c_n = b_1c_1 + \sum_{n=2}^{k}(\Delta B_{n-1})c_n = b_1c_1 + \sum_{n=2}^{k}(\Delta B_{n-1})d_{n-1}$$

$$= B_1c_1 + \sum_{n=2}^{k}(\Delta B_{n-1})\,d_{n-1} = B_1c_1 + \left[B_{k+1}d_{k+1} - B_2d_2\right] - \sum_{n=2}^{k} B_{n+1}\Delta d_n \quad (1.14)$$

$$= B_1c_1 + \left[B_{k+1}c_{k+2} - B_2c_3\right] - \sum_{n=2}^{k} B_{n+1}\Delta c_{n+1}\,. \quad (1.15)$$

Dunque, per garantire la convergenza della serie $\sum_{n=1}^{+\infty} b_n c_n$, basta dare condizioni che garantiscano l'esistenza dei due limiti

$$\lim_{k\to+\infty} B_k c_{k+1}\,, \tag{1.16}$$

$$\lim_{k\to+\infty} \left[\sum_{n=2}^{k} B_{n+1}\Delta c_{n+1}\right]\,. \tag{1.17}$$

Un criterio per questo è dato dal *Teorema di Dirichlet*, che immediatamente implica il criterio di Leibniz:

Teorema 42 (*Teorema di Dirichlet*) *Sia* $\{c_n\}_{n\ge1}$ *una successione* a valori positivi, decrescente e convergente a zero. *Sia* $\sum_{n=1}^{+\infty} b_n$ *una serie tale che* $\{B_n\}_{n\ge1}$ rimane limitata. *Allora, la serie* $\sum_{n=1}^{+\infty} b_n c_n$ *converge.*

Dim. Infatti, il limite (1.16) è nullo perché $\{B_n\}$ è limitata e $c_n \to 0$. Il limite (1.17) esiste per il Lemma 40. ∎

Dimostrazione del criterio di convergenza di Leibniz. La dimostrazione del criterio di Leibniz per la serie a segni alterni

$$\sum_n (-1)^n a_n\,, \qquad a_n \ge 0$$

con $\{a_n\}$ decrescente e convergente a zero, segue immediatamente: basta definire $b_n = (-1)^n$ e $c_n = a_n \to 0$ e notare che

$$B_n = \begin{cases} -1 & \text{se } n \text{ è dispari} \\ +1 & \text{se } n \text{ è pari} \end{cases}$$

e quindi $\{B_n\}_{n\ge1}$ rimane limitata.

Una diversa condizione per l'esistenza dei due limiti (1.16) e (1.17) è data dal teorema seguente:

Teorema 43 (*Teorema di Abel*) *Sia* $\{c_n\}_{n\ge1}$ *una successione* a valori positivi e decrescente. *Sia* $\sum_{n=1}^{+\infty} b_n$ *una serie convergente. Allora, la serie* $\sum_{n=1}^{+\infty} b_n c_n$ *converge.*

Dim. Per ipotesi, esistono ambedue i limiti $\lim c_k$ e $\lim B_k$. Dunque, esiste il limite (1.16). La successione $\{B_k\}$, essendo convergente è anche limitata e quindi il limite (1.17) esiste per il Lemma 40. ∎

Capitolo 2

Successioni e serie di funzioni

In questo capitolo studiamo le successioni e le serie di funzioni. Studiamo quindi particolari metodi per approssimare una data funzione $f(x)$ mediante serie i cui termini sono funzioni elementari: monomi o funzioni trigonometriche.

2.1 Introduzione

Ricordiamo, dal Par. 1.5.1, che una *successione di funzioni* è una trasformazione che ad ogni numero naturale n associa una funzione $f_n(x)$ definita su un dominio I, lo stesso per ogni n. Stando così le cose, il problema della convergenza della successione di funzioni si studia per ogni fissato $x \in I$; ossia si studia una successione numerica per ogni valore del parametro x. Se la succesisone $(f_n(x))$ converge a $f(x)$ per ogni $x \in I_0 \subseteq I$, si dice che la successione $(f_n(x))$ *converge puntualmente* ad $f(x)$ su I_0.

In realtà questo è un modo piuttosto semplicistico di procedere. Le applicazioni richiedono infatti di studiare la convergenza rispetto ad opportuni concetti di "distanza" tra funzioni. Ciò è quanto facciamo in questo capitolo.

Notazione. Da ora in poi di regola useremo una notazione meno "elementare": per indicare una funzione invece di scrivere $f(x)$ scriveremo semplicemente f. Invece, col simbolo $f(x)$ intenderemo il valore che la funzione f assume nel punto x. In certi casi questo può condurre ad ambiguità ed allora useremo notazioni del tipo $x \to f(x)$ per indicare la funzione che ad x associa $f(x)$. Va detto che la notazione f non si può usare per specifiche funzioni: la funzione $x \to \sin x$ si indicherà semplicemente con $\sin x$. Dunque, successioni e serie di funzioni di regola si indicheranno con la notazione

$$(f_n)\,, \qquad \sum f_n\,, \qquad \sum_{n=1}^{+\infty} f_n$$

ecc. Invece scriveremo esplicitamente x quando dovremo considerare successioni o serie di funzioni particolari, per esempio

$$\sum 5^n(x-1)^n\,, \qquad \sum \frac{1}{nx^2+1} \qquad \sum \frac{1}{n^2}\sin nx\,. \quad \blacksquare$$

2.2 Distanze tra funzioni

Noi considereremo funzioni definite su un fissato insieme I che sarà sempre un intervallo. Specificheremo volta per volta se tale intervallo deve essere limitato e/o chiuso. Inoltre, non considereremo tutte le possibili funzioni definite su I, ma le funzioni di una delle tre classi seguenti:

- col simbolo $C(I)$ si intende l'insieme delle funzioni continue su I. Implicitamente usando questo simbolo si intende che I sia limitato e chiuso, $I = [a, b]$.
- col simbolo $\mathcal{L}^1(I)$ si intende l'insieme delle funzioni il cui valore assoluto è integrabile su I:
$$\int_I |f(x)|\,\mathrm{d}x < +\infty\,.$$
L'insieme I è ora un intervallo qualsiasi.
- col simbolo $\mathcal{L}^2(I)$, si intende l'insieme delle funzioni tali che
$$\int_I |f(x)|^2\,\mathrm{d}x < +\infty\,.$$
L'insieme I è ora un intervallo qualsiasi.

  Si noti la presenza del segno $|\cdot|$ e del quadrato. Ovviamente, se la funzione prende valori reali, $|f(x)|^2 = f^2(x)$. Però molto spesso la funzione $f(x)$ prende valori complessi e in tal caso il segno di modulo è necessario.

Osservazione 44 L'integrale che usiamo è l'integrale di Riemann. Si ricordi che una funzione integrabile secondo Riemann è necessariamente limitata. Usando questa proprietà, si potrebbe mostrare che, se l'intervallo I è limitato, i due insiemi $\mathcal{L}(I)$ ed $\mathcal{L}^2(I)$ hanno i medesimi elementi. I due insiemi invece sono diversi se I è illimitato o se si usano concetti diversi di integrale, che permettono di integrare anche funzioni illimitate.

A noi interessa sapere che:
$$I \text{ limitato } \Rightarrow \mathcal{L}^2(I) \subseteq \mathcal{L}^1(I)\,. \quad \blacksquare$$

I tre insiemi sopra definiti sono spazi lineari, nel senso visto nei corsi di geometria, rispetto alle operazioni
$$(f+g)(x) = f(x) + g(x)\,, \qquad (\alpha f)(x) = \alpha f(x)\,.$$

In questi spazi lineari si introduce una *distanza* come segue:

- in $C(I)$ si pone
$$d(f,g) = \max_{x\in I} |f(x) - g(x)|$$
(si noti che il massimo esiste, dal teorema di Weierstrass, perché l'intervallo I è limitato e chiuso.) La distanza sopra definita si indica anche col simbolo[1]
$$d_\infty(f,g)\,.$$

[1] non spieghiamo la ragione dell'indice "∞". Per spiegarlo, dovremmo introdurre una famiglia di altre distanze, che non useremo.

- Le distanze in $\mathcal{L}^1(I)$ e in $\mathcal{L}^2(I)$ si indicano anche, rispettivamente, con i simboli

$$d_1(f,g)\,,\qquad d_2(f,g)$$

e sono definite da

$$d_1(f,g)=\int_I |f(x)-g(x)|\,\mathrm{d}x\,,\qquad d_2(f,g)=\left[\int_I |f(x)-g(x)|^2\,\mathrm{d}x\right]^{1/2}.$$

Si noti che in questi simboli l'intervallo I è sottinteso e che spesso si scrive, rispettivamente,

$$||f-g||_\infty\,,\qquad ||f-g||_1\,,\qquad ||f-g||_2\,.$$

usando il simbolo di "norma della differenza", al posto di quello di distanza

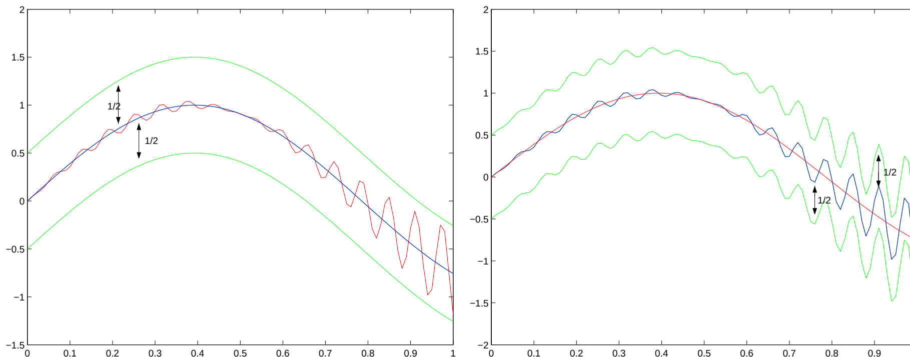
Osservazione 45 Non ci serve essere precisi su questo. Però va detto che solamente la norma $||\cdot||_\infty$ è effettivamente una norma e che solo $d_\infty(f,g)$ è effettivamente una distanza. Infatti, se $I=[0,1]$, se $f(x)=0$ per ogni x e se $g(x)=0$ per $x\neq 1/2$ mentre $g(1/2)=1$, allora $d_1(f,g)=0$, $d_2(f,g)=0$ nonostante che f e g siano tra loro diverse. ∎

Accade che

$$d_\infty(f,g)<\epsilon$$

quando il grafico di $g(x)$ sta in un "tubo" di ampiezza ϵ intorno a quello di $f(x)$, come in figura 2.1. Si noti che il "tubo" può essere disegnato intorno ad una qualsiasi delle due funzioni. Ciò fatto si richiederà che il grafico dell'altra stia in tale "tubo".

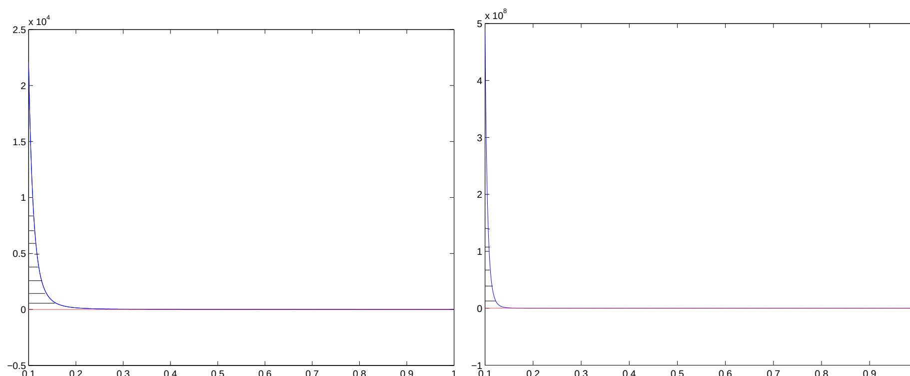
Figura 2.1: Distanza $d_\infty(f,g)$



Invece, $d_1(f,g)<\epsilon$ quando accade che è minore di ϵ l'area del trapezoide di $|f(x)-g(x)|$, ossia l'area della superficie compresa tra i grafici delle due funzioni.

La figura 2.2, a sinistra, riporta il grafico di $|f(x)-g(x)|$ ed a destra quello di $|f(x)-g(x)|^2$ (per le medesime due funzioni $f(x)$ e $g(x)$) con tratteggiata la regione la cui area deve essere "piccola" se si vuole che queste "distino di poco" rispettivamente in $\mathcal{L}^1(a,b)$ ed

Figura 2.2: Le distanze $d_1(f,g)$ e $d_2(f,g)$



in $\mathcal{L}^2(a,b)$. La figura suggerisce che la distanza, nel senso $\mathcal{L}^1$, o $\mathcal{L}^2$, tra due funzioni possa essere "piccola" anche se i due grafici diventano "molto distanti" per certi valori di x.

Scriviamo ora esplicitamente la definizione di convergenza rispetto alla distanza d_∞. Questo tipo di convergenza si chiama *convergenza uniforme*.

La successione di funzioni $(f_n(x))$ *converge uniformemente* ad $f(x)$ su I se per ogni $\epsilon > 0$ esiste N_ϵ tale che per ogni $n > N_\epsilon$ vale

$$\max_{x \in I} |f_n(x) - f(x)| < \epsilon \qquad \boxed{\text{per ogni } x \in I,}$$

ossia, per $n > N_\epsilon$ vale

$$|f_n(x) - f(x)| < \epsilon \qquad \boxed{\text{per ogni } x \in I.}$$

Per contrasto, diamo la definizione di convergenza puntuale: la successione $(f_n(x))$ converge puntualmente ad $f(x)$ su I quando per ogni $x \in I$ e per ogni $\epsilon > 0$ esiste un numero $N_\epsilon(x)$ tale che se $n > N_\epsilon(x)$ si ha, per questo particolare numero x, $|f_n(x) - f(x)| < \epsilon$. Cambiando il punto x cambia anche il numero $N_\epsilon(x)$. Si ha convergenza uniforme quando si può trovare un numero N_ϵ che va bene per ogni $x \in I$.

Diamo ora la definizione di convergenza in $\mathcal{L}^1(I)$ ed in $\mathcal{L}^2(I)$.

La successione $(f_n(x))$ converge ad $f(x)$ in $\mathcal{L}^1(I)$ se per ogni $\epsilon > 0$ esiste N_ϵ tale che per ogni $n > N_\epsilon$ si ha $$\int_I |f_n(x) - f(x)|\,\mathrm{d}x < \epsilon\,.$$	La successione $(f_n(x))$ converge ad $f(x)$ in $\mathcal{L}^2(I)$ se per ogni $\epsilon > 0$ esiste N_ϵ tale che per ogni $n > N_\epsilon$ si ha $$\int_I |f_n(x) - f(x)|^2\,\mathrm{d}x < \epsilon^2\,.$$

La convergenza in $\mathcal{L}^1(I)$ si chiama anche *convergenza in media* e la convergenza in $\mathcal{L}^2(I)$ si chiama anche *convergenza in* *media quadratica* mentre, come si è già detto, la convergenza in $C(I)$ si chiama *convergenza uniforme.*

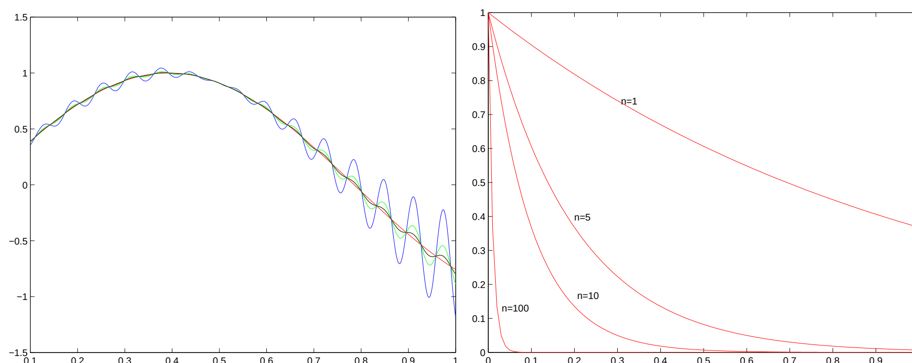
Osservazione 46 Supponiamo che la successione di funzioni (f_n) sia definita su un dominio I, $I = I_1 \cup I_2$. Se la successione converge uniformemente su I, essa converge uniformemente sia su I_1 che su I_2 e, viceversa, se la successione converge uniformemente su I_1 e anche su I_2, essa converge uniformemente anche su I. Asserto analogo vale anche per la convergenze in media e per la convergenze in media quadratica. ∎

Vediamo ora due figure che illustrano il significato della convergenza uniforme e della convergenza in media. La fig. 2.3 a sinistra suggerisce la convergenza uniforme. La figura a destra vuol suggerire la convergenza in media, e vuol sottolineare che la convergenza in media non implica la convergenza puntuale.

I grafici rappresentati a destra sono quelli delle funzioni e^{-nt} con $n = 1, 5, 10, 100$.

Si provi dettagliatamente, facendo uso della definizione, che la successione (e^{-nt}), $t \in [0, 1]$, converge a zero in media ed in media quadratica, ma non uniformemente.

Figura 2.3: Convergenza uniforme e convergenza in media



Infine, enunciamo il lemma seguente, immediata conseguenza del teorema di confronto per i limiti. La dimostrazione è lasciata per esercizio.

Lemma 47 *Sia (f_n) una successione di funzioni su un intervallo I. Su tale intervallo sia anche definita una funzione g. Sia (γ_n) una successione numerica tale che*

$$\lim \gamma_n = 0\,.$$

- *se vale*

$$d_\infty(f_n, g) < \gamma_n$$

  allora la successione (f_n) converge uniformemente a g.

- *se vale*
$$d_1(f_n, g) < \gamma_n$$
allora la successione (f_n) *converge a* g *in* $\mathcal{L}^1(I)$.
- *se vale*
$$d_2(f_n, g) < \gamma_n$$
allora la successione (f_n) *converge a* g *in* $\mathcal{L}^2(I)$.

2.2.1 Il prodotto interno su $\mathcal{L}^2(a, b)$

Su $\mathcal{L}^2(a, b)$ si può definire un *prodotto interno* o *prodotto integrale* come segue: siano f e g due funzioni a quadrato integrabile. Si può provare che il loro prodotto è integrabile. Definiamo allora il prodotto interno delle due funzioni f e g ponendo[2]

$$\langle f, g\rangle = \int_a^b \bar{g}(s) f(s)\, \mathrm{d}s\,.$$

Si noti che se le funzioni prendono valori reali allora il segno di coniugio non ha alcun effetto; se però esse prendono valori complessi il coniugio è importante perché è grazie ad esso che si ottiene

$$\sqrt{\langle f, f\rangle} = ||f||_2\,.$$

Quest'uguaglianza giustifica il termine "prodotto interno" (o "prodotto scalare"): nel caso dei "vettori ordinari" di $\mathbf{R}^3$ si ha infatti

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2} = \mathbf{x} \cdot \mathbf{x}\,.$$

Diciamo che due funzioni f e g sono *ortogonali* in $\mathcal{L}^2(a, b)$ quando

$$\langle f, g\rangle = 0\,.$$

Naturalmente, per dire che f è ortogonale a g, scriveremo

$$f \perp g\,.$$

Una proprietà importante del prodotto interno in $\mathcal{L}^2(a, b)$ è che per esso vale il *teorema di Pitagora:*

Teorema 48 *Se* $f \perp g$ *allora* $||f + g||^2 = ||f||^2 + ||g||^2$.

Dim. Infatti si ha

$$||f + g||^2 = \langle f + g, f + g\rangle = \langle f, f\rangle + \langle f, g\rangle + \langle g, f\rangle + \langle g, g\rangle = ||f||^2 + ||g||^2\,. \quad \blacksquare$$

In particolare,

$$f \perp g \Longrightarrow ||f|| \le ||f + g||\,, \quad ||g|| \le ||f + g||\,.$$

Osservazione 49 Va notato esplicitamente che non esiste nessun concetto analogo a quello di prodotto interno per le distanze in $C(a, b)$ o di $\mathcal{L}^1(a, b)$. ■

---

[2]si può mostrare che le proprietà essenziali di questo prodotto mimano quelle del prodotto scalare di vettori di $\mathbf{R}^n$ o di $\mathbf{C}^n$.

2.2.2 Proprietà della convergenza uniforme

La sola convergenza puntuale è troppo debole per avere proprietà importanti di qualche tipo. Infatti:

- esistono successioni di funzioni integrabili che convergono puntualmente a funzioni non integrabili.

Per vedere un esempio, ricordiamo che i numeri razionali sono numerabili, ossia sono immagine di una successione iniettiva. Sia (q_k) questa successione e definiamo la funzione

$$f_n(x) = \begin{cases} 1 & \text{se} \quad x = q_1\,, q_2\,, \dots, q_n \\ 0 & \quad\text{altrimenti}\,. \end{cases}$$

Ciascuna delle funzioni $f_n(x)$ è integrabile, con integrale uguale a zero, perché $f_n(x)$ è zero salvo che in un numero finito di punti.

La successione $(f_n(x))$ converge alla *funzione di Dirichlet,* che vale 1 sui razionali e vale 0 sugli irrazionali. Questa funzione non è integrabile.

- esistono successioni di funzioni continue che convergono puntualmente a funzioni discontinue.

Un esempio è il seguente: sia $I = [0, 1]$ e sia $f_n(x) = x^n$. Ciascuna funzione è continua, ma la successione di funzioni $(f_n(x))$ converge puntualmente su I alla funzione

$$f(x) = \begin{cases} 0 & \text{se} \quad x \in [0, 1) \\ 1 & \text{se} \quad x = 1\,. \end{cases}$$

Questa funzione è discontinua.

E' per questa ragione che vanno introdotti concetti più restrittivi di convergenza. In questo paragrafo presentiamo le relazioni tra il concetto di convergenza uniforme e le proprietà di limite, derivabilità ed integrazione[3].

Il teorema centrale è il seguente:

Teorema 50 *Sia $(f_n(x))$ una successione di funzioni, definite su un intervallo I, ciascuna delle quali è continua in un punto x_0. Se la successione converge ad una funzione $f(x)$* uniformemente su I *allora anche la funzione $f(x)$ è continua in x_0.*

Omettiamo la dimostrazione.

Ricordando che le funzioni continue su un intervallo limitato e chiuso sono integrabili si ha:

Corollario 51 *Sia $(f_n(x))$ una successione di funzioni continue su un intervallo limitato e chiuso I. La successione converga uniformemente su I ad una funzione $f(x)$. Allora, la funzione $f(x)$ è continua e quindi anche integrabile su I.*

Sotto le ipotesi del corollario precedente, può dirsi di più: l'integrale di $f(x)$ è il limite degli integrali delle $f_n(x)$. Ossia:

[3]invece, non illustreremo le relazioni tra tali proprietà e le convergenze in media, che richiederebbero l'introduzione di concetti di integrale più generali di quello di Riemann.

Teorema 52 *Sia $(f_n(x))$ una successione di funzioni continue su un intervallo limitato e chiuso I. La successione converga uniformemente su I ad una funzione $f(x)$. Allora,*

$$\int_I f(x)\, dx = \int_I \left[\lim_{n\to+\infty} f_n(x)\right]\, dx = \lim_{n\to+\infty}\left[\int_I f_n(x)\, dx\right]. \tag{2.1}$$

Dim. Ricordando la definizione di limite, dobbiamo provare che per ogni $\epsilon > 0$ esiste N_ϵ tale che per ogni $n > N_\epsilon$ si ha

$$\left|\int_I f(x)\, \mathrm{d}x - \int_I f_n(x)\, \mathrm{d}x\right| < \epsilon\,.$$

Dalla linearità e dalla proprietà di monotonia degli integrali,

$$\left|\int_I f(x)\, \mathrm{d}x - \int_I f_n(x)\, \mathrm{d}x\right| \le \int_I |f_n(x) - f(x)|\, \mathrm{d}x$$

e quindi basta provare che

$$\int_I |f_n(x) - f(x)|\, \mathrm{d}x < \epsilon$$

per n sufficientemente grande.

Ricordiamo che la successione $(f_n(x))$ converge ad $f(x)$ uniformemente su I: scelto un qualunque $\sigma > 0$ esiste $\tilde{N}_\sigma$ tale che, per ogni $n > \tilde{N}_\sigma$ si ha

$$|f_n(x) - f(x)| < \sigma \qquad \forall x \in I$$

e quindi anche

$$\int_I |f_n(x) - f(x)|\, \mathrm{d}x < \sigma L$$

con L la lunghezza di I. L'asserto segue scegliendo $\sigma = \epsilon/L$ e $N_\epsilon = \tilde{N}_{\epsilon/L}$. ∎

Osservazione 53 Ricordiamo, dall'Osservazione 46, che se $I = I_1 \cup I_2$ e se la successione (f_n) converge uniformemente su I, essa converge uniformemente sia su I_1 che su I_2. D'altra parte, l'integrale su $I = I_1 \cup I_2$ è la somma dei due integrali su I_1 e su I_2. Quindi il teorema si applica facilmente anche a successioni uniformemente convergenti di funzioni continue a tratti e questo è sufficiente per molte applicazioni. Si potrebbe però provare che la semplice integrabilità di ciascuna delle funzioni $f_n(x)$ e la convergenza uniforme della successione ad $f(x)$ implica che la $f(x)$ è integrabile e che vale l'uguaglianza (2.1). ∎

Sia $F_n(x)$ primitiva di $f_n(x)$. La convergenza della successione (f_n) non implica la convergenza della successione (F_n). Questo si vede facilmente scegliendo $f_n(x) = 0$ per ogni x, ed $F_n(x) = n$ per ogni x. Le cose cambiano se si assegna il comportamento che le $F_n(x)$ devono avere in un medesimo punto x_0:

Teorema 54 *Sia $(f_n(x))$ una successione di funzioni continue su un intervallo $[a, b]$ e sia*

$$F_n(x) = c_n + \int_a^x f_n(s)\, ds\,.$$

Supponiamo che

$$\lim_n c_n = c_0 \,,$$

e che la successione (f_n) converga uniformemente *ad* f_0. *In tal caso la successione* (F_n) converge uniformemente *alla funzione*

$$F_0(x) = c_0 + \int_a^x f_0(s)\, ds \,.$$

Dim. Ricordiamo che la funzione $f_0(x)$ è continua, come limite uniforme di una successione di funzioni continue. Dunque l'espressione proposta per $F_0(x)$ ha senso.

La differenza $|F_n(x) - F_0(x)|$ si stima così:

$$|F_n(x) - F_0(x)| \le |c_n - c_0| + \int_a^x |f_n(s) - f_0(s)| \, \mathrm{d}s \le |c_n - c_0| + \int_a^b |f_n(s) - f_0(s)| \, \mathrm{d}s \,.$$

La successione numerica $(|c_n - c_0|)$ tende a zero per ipotesi, mentre la successione numerica

$$\left\{ \int_a^b |f_n(s) - f_0(s)| \, \mathrm{d}s \right\}$$

tende a zero per il Teorema 52. L'asserto segue quindi dal Lemma 47. ∎

Vediamo ora le relazioni tra derivazione e convergenza uniforme. E' facile trovare esempi che mostrano che il limite uniforme di una successione di funzioni derivabili può non essere derivabile.

Esempio 55 Consideriamo la successione di funzioni

$$f_n(x) = \begin{cases} |x| & \text{se} \quad |x| > 1/n \\ \frac{n}{2}x^2 + \frac{1}{2n} & \qquad \text{altrimenti} \,. \end{cases}$$

Ciascuna di queste funzioni è derivabile ma la successione di funzioni converge uniformemente alla funzione non derivabile $f(x) = |x|$. Il grafico di alcune di queste funzioni è in figura 2.4. ∎

Vale però:

Teorema 56 *Sia* (f_n) *una successione di funzioni derivabili su un intervallo* (a, b). *Supponiamo che:*

- *esista un punto* x_0 *tale che la successione numerica* $(f_n(x_0))$ *converga;*
- *la successione* $(f'_n(x))$ converga uniformemente *su* I *ad una funzione* $h(x)$.

In tal caso la successione (f_n) converge uniformemente su I *ad una funzione* $f(x)$ *che è derivabile, ed inoltre* $f'(x) = h(x)$*; ossia*

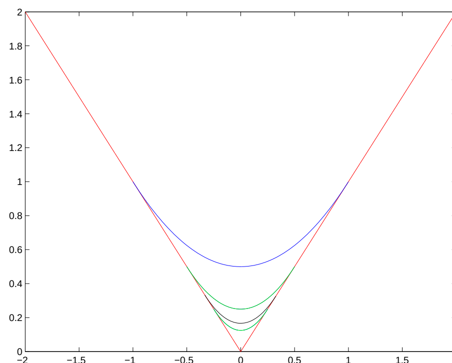
$$f'(x) = \lim_{n \to +\infty} f'_n(x) \,.$$

Dim. Per ipotesi,

$$f_n(x) = f_n(x_0) + \int_{x_0}^x f'_n(s) \, \mathrm{d}s \,.$$

L'asserto segue quindi dal Teorema 54. ∎

Figura 2.4: La convergenza uniforme non implica la convergenza della successione delle derivate



2.3 Serie di funzioni

Le *serie di funzioni* si definiscono in modo del tutto analogo alle serie numeriche: se (f_n) é una successione di funzioni, la serie corrispondente si indica col simbolo

$$\sum_n f_n \qquad \text{oppure} \qquad \sum_n f_n(x)\,.$$

Se é necessario indicare i valori presi dall'indice scriveremo per esempio

$$\sum_{n=n_0}^{+\infty} f_n\,.$$

Il significato del simbolo di serie é il seguente: supponendo che l'indice prenda tutti i valori maggiori od uguali a 0, si costruisce la nuova successione

$$s_0(x) = f_0(x)\,, \quad s_1(x) = f_0(x) + f_1(x)\,, \quad s_n(x) = f_0(x) + f_1(x) + \cdots + f_n(x)\,.$$

Se la successione di funzioni $(s_n(x))$ converge uniformemente, si dice che *la serie converge uniformemente;* se la successione di funzioni $(s_n(x))$ converge in $\mathcal{L}^1$ oppure in $\mathcal{L}^2$, si dice che la serie converge in tal senso.

Enunciamo il teorema seguente, senza dimostrarlo:

Teorema 57 *Sia convergente la serie di numeri* positivi

$$\sum \gamma_n\,.$$

Valga inoltre

$$d_p(f_n, 0) \le \gamma_n$$

ove $p = +\infty$ oppure $p = 1$oppure $p = 2$. Sotto queste ipotesi, esiste una funzione $g(x)$ (rispettivamente in $C(I)$, $\mathcal{L}^1(I)$ oppure $\mathcal{L}^2(I)$) tale che la serie

$$\sum f_n$$

converge nel senso corrispondente a $g(x)$.

Sia $\sum f_n$ una serie di funzioni. Si dice che questa serie *converge normalmente* (in $C(I)$ oppure in $\mathcal{L}^p(I)$) quando converge la serie

$$\sum ||f_n||_p$$

(con p uguale a 1, 2). L'asserto del teorema precedente si esprime anche dicendo che una serie che converge normalmente è anche una serie convergente. Il viceversa non vale.

Nel caso di $p = +\infty$, ossia per la convergenza uniforme, il test per la convergenza dato nel Teorema 57 si chiama anche *test di Weierstrass*.

A parte questo teorema, di dimostrazione difficile, i teoremi visti per la convergenza delle successioni di funzioni si adattano facilmente allo studio della convergenza delle serie. Enunciamo in particolare:

Teorema 58 *Una serie di funzioni* continue $\sum f_n$ *converga uniformemente su I ad una funzione $f(x)$. Allora la funzione $f(x)$ è* continua *e, se I è limitato e chiuso, vale*

$$\int_I \left[\sum_n f_n(x)\right] dx = \sum_n \left[\int_I f_n(x)\, dx\right] .$$

Sia $\sum_n f_n(x)$ una successione di funzioni derivabili, convergente puntualmente *su I ad una funzione $f(x)$. Supponiamo che la serie $\sum_n f_n'(x)$ converga* uniformemente *su I ad una funzione $g(x)$. Allora la funzione $f(x)$ è derivabile, con derivata $f'(x) = g(x)$ ed inoltre la serie $\sum_n f_n(x)$ converge ad $f(x)$* uniformemente *su I.*

Esempio 59 Consideriamo la serie

$$\sum_{n=0}^{+\infty} x(1-x)^n\,, \qquad x \in [0,1]\,.$$

Questa serie converge puntualmente su $[0,1]$. ossia converge per ogni $x \in [0,1]$. Infatti, se $x = 0$ ogni termine della serie è nullo e quindi la somma della serie è 0; se $x \in (0,1]$ allora è lecito scrivere

$$\sum_{n=0}^{+\infty} x(1-x)^n = x\left(\sum_{n=0}^{+\infty}(1-x)^n\right) = x\frac{1}{1-(1-x)} = 1\,.$$

Dunque, si ha

$$\sum_{n=0}^{+\infty} x(1-x)^n = \begin{cases} 0 & \text{se } x = 0 \\ 1 & \text{altrimenti.} \end{cases}$$

Dunque la somma della serie non è continua e quindi la serie, pur convergendo puntualmente su $[0,1]$, non converge uniformemente. ∎

2.4 Serie di potenze

Si chiamano serie di potenze le serie di funzioni della forma

$$\sum_{n=0}^{+\infty} a_n[x-x_0]^n \,, \tag{2.2}$$

ottenute a partire dalla successione di monomi $a_n[x-x_0]^n$.

Si noti che il primo valore dell'indice è ora $n=0$ ma niente vieta che sia $a_0=0$ ossia che la serie "parta" per esempio con indice 1.

Se esiste N tale che $a_n=0$ per $n>N$, allora la serie di potenze si riduce ad un polinomio: le serie di potenze generalizzano i polinomi.

Il numero x_0 si chiama il *centro* della serie e la serie di potenze converge sempre per $x=x_0$ (e ivi converge ad a_0). Potrebbe non convergere in nessun altro punto.

Esempio 60 Si consideri la serie seguente, di centro $x_0=0$:

$$\sum_{n=0}^{+\infty} n^n x^n = \sum_{n=0}^{+\infty} (nx)^n \,.$$

Si fissi il valore di $x\neq 0$ e sia n_0 tale che $|n_0 x|>1$. Allora, per $n>n_0$, si ha

$$|nx|^n > |n_0 x|^n \to +\infty \,.$$

Dunque, se $x\neq 0$, il termine generale della serie non tende a zero, e quindi la serie non converge. ∎

Vale però:

Teorema 61 (Teorema di Abel) *Una serie di potenze converga in un punto $\xi\neq x_0$ e sia*

$$r=|\xi-x_0|>0\,.$$

Allora, la serie di potenze converge uniformemente in $\{x \mid |x-x_0|<r'\}$ per ogni $r'<r$.

Dim. Si guardi la figura 2.5 per seguire questa dimostrazione.

Sia $|x-x_0|<r'<r$ e sia $d\in(r',r)$ così che

$$r'<r\,, \qquad r=|x-\xi|>d$$

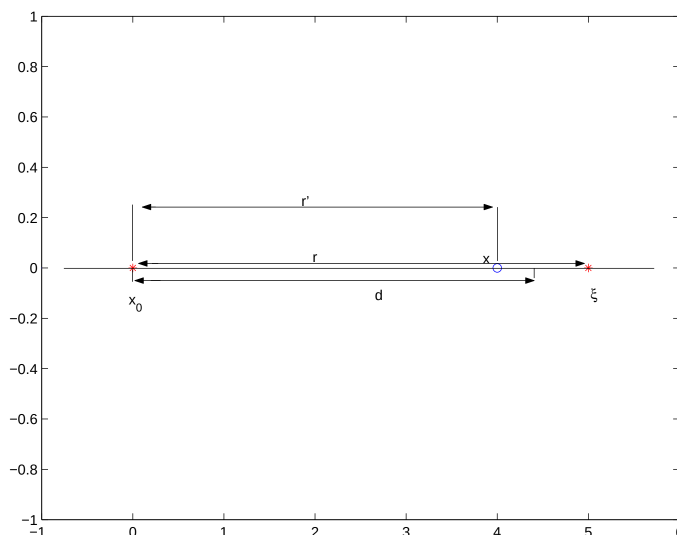
e quindi

$$|a_n(x-x_0)^n| < |a_n(\xi-x_0)^n|\cdot\left(\frac{r'}{|\xi-x_0|}\right)^n \le |a_n(\xi-x_0)^n|\cdot\left(\frac{r'}{d}\right)^n \,.$$

La convergenza in ξ implica che la successione $(|a_n(\xi-x_0)^n|)$ è limitata (anzi tende a zero):

$$|a_n(\xi-x_0)^n|<M\,.$$

Figura 2.5: Gli intervalli che si usano nella dimostrazione del teorema di Abel



Inoltre,

$$\frac{r'}{d} < q < 1 .$$

Dunque, per $|x - x_0| < r'$ si ha

$$|a_n(x - x_0)^n| \le \gamma_n , \qquad \gamma_n \le Mq^n$$

con $0 \le q < 1$ e quindi con $\sum \gamma_n < +\infty$. La convergenza uniforme segue dal criterio di Weierstrass, Teorema 57. ∎

Questo risultato in particolare implica che se una serie di potenze converge allora l'insieme su cui essa converge è un intervallo centrato in x_0 (e non si esclude che sia ridotto al solo x_0, oppure che sia tutta la retta). Questo si chiama l' *intervallo di convergenza* della serie di potenze e si chiama *raggio di convergenza* la sua semiampiezza. Il raggio di convergenza può essere nullo, si veda l'esempio 60. Può essere un numero finito non nullo oppure può essere $+\infty$: è un numero finito non nullo nel caso particolare della serie geometrica, è $+\infty$ nel caso in cui la serie di potenze si riduca ad un polinomio, ma non solo in tal caso:

Osservazione 62 Consideriamo la serie

$$\sum_{n=0}^{+\infty} \frac{x^n}{n!} .$$

Fissato x, si trova una serie numerica, convergente per ogni valore del parametro x, come si vede facilmente usando il criterio del rapporto. ∎

Inoltre, la dimostrazione del Teorema 61 mostra che:

Teorema 63 (Teorema di Abel) *Sia $0 \leq R \leq +\infty$ il raggio di convergenza di una serie di potenze e sia x tale che $|x - x_0| < R$. La serie di potenze converge assolutamente e quindi incondizionatamente nel punto x.*

Ciò mostra che l'ordine dei termini di una serie di potenze non influisce sulla somma della serie nei punti interni all'intervallo di convergenza. Invece, niente può dirsi in generale del comportamento della serie negli estremi dell'intervallo di convergenza, come mostrano gli esempi seguenti.

Esempio 64 Consideriamo la serie di potenze

$$\sum_{n=1}^{+\infty} (-1)^2 \frac{x^n}{n} \tag{2.3}$$

Fissato x si trova una serie numerica e si vede facilmente, dal criterio del rapporto, che la sere converge per ogni x con $|x| < 1$, diverge se $|x| > 1$. Dunque il raggio di convergenza è 1. Se $x = -1$ oppure $x = +1$ si trovano rispettivamente le due serie

$$\sum_{n=1}^{+\infty} \frac{1}{n}\,, \qquad \sum_{n=1}^{+\infty} (-1)^n \frac{1}{n}\,.$$

La prima divergente e la seconda convergente.

Si è detto che la somma di una serie di potenze non dipende dall'ordine degli elementi nei punti interni dell'intervallo di convergenza. Quest'esempio mostra che, invece, può dipendere dall'ordine degli elementi se x è uno degli estremi dell'intervallo di convergenza.

Mostriamo ora una serie di potenze che converge in ambedue gli estremi dell'intervallo di convergenza. La serie è

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} x^{2n}\,. \tag{2.4}$$

Di nuovo, fissato x, si usi il criterio del rapporto per studiare la convergenza della serie numerica che si ottiene. Si vede che si ha convergenza per $-1 \leq x \leq 1$, estremi inclusi, mentre la serie diverge per $|x| > 1$.

Infine, mostriamo una serie di potenze che diverge in ambedue gli estremi dell'intervallo di convergenza. Questa è la serie

$$\sum_{n=0}^{+\infty} (-1)^n x^{2n}\,. \tag{2.5}$$

Questa è una serie geometrica di ragione $-x^2$ e quindi converge se e solo se $-1 < x < 1$, estremi esclusi.

Si noti che la somma di questa serie è la funzione

$$f(x) = \frac{1}{1 + x^2}\,,$$

una funzione di classe $C^\infty(R)$. La ragione per cui la serie corrispondente non debba convergere negli estremi sembra misteriosa, e verrà chiarita al paragrafo 2.4.2.

E' interessante vedere i grafici di alcune delle somme parziali delle tre serie introdotte sopra. L'intervallo di convergenza delle serie è $(-1, 1)$. I grafici delle somme parziali sono disegnati su un intervallo un po' più grande e, punteggiato, si riporta anche il grafico della somma della serie, ovviamente solo sull'intervallo $(-1, 1)$. La prima e la seconda serie sono alla figura 2.6 L'ultima serie è alla figura 2.7. Il grafico punteggiato (tratto spesso) in

Figura 2.6: Somme parziali dellaserie (2.3) (a sinistra e della serie (2.4) (a destra)

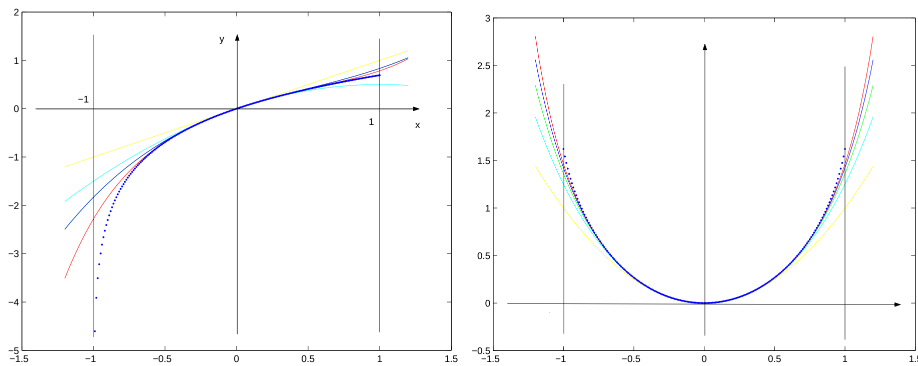
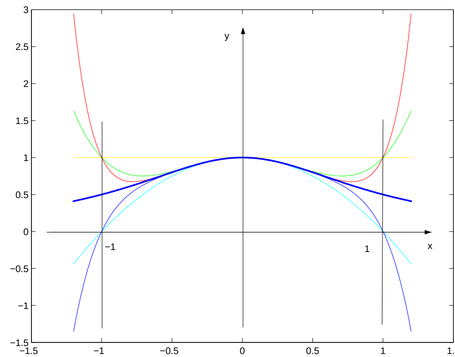


Figura 2.7: Somma parziali della serie (2.5)



questo caso rappresenta la funzione $1/(1+x^2)$, che è la somma della serie sul solo intervallo $(-1, 1)$. E' stato disegnato in un intervallo più grande per sottolineare che la questa funzione è regolare anche nei punti $+1$ e -1. ∎

Inoltre, la convergenza essendo uniforme, dal Teorema 50:

Corollario 65 *La somma di una serie di potenze è continua nei punti interni all'intervallo di convergenza.*

Si potrebbe inoltre provare che:

Teorema 66 *Sia R il raggio di convergenza della serie $\sum a_n(x-x_0)^n$. Le due serie di potenze*

$$\sum \frac{d}{dx}\left[a_n(x-x_0)^n\right] = \sum n a_n (x-x_0)^{n-1}\,,$$
$$\sum \left[\int_{x_0}^{x} a_n(s-x_0)^n\, ds\right] = \sum \frac{1}{n+1} a_n (x-x_0)^{n+1} \qquad (2.6)$$

hanno lo stesso raggio di convergenza e questo è uguale ad R.

Le due serie in (2.6) si dicono ottenute *derivando* *termine a termine* o *integrando* *termine a termine* la serie data. Esse convergono uniformemente in $[a,b] \subseteq (x_0-R, x_0+R)$ e quindi, usando i teoremi 52 e 56, si trova l'uguaglianza

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\sum a_n(x-x_0)^n\right] = \sum n a_n (x-x_0)^{n-1}\,,$$
$$\int_0^x \left[\sum a_n(x-x_0)^n\right] = \sum \frac{1}{n+1} a_n (x-x_0)^{n+1}\,.$$

In particolare quindi anche la serie derivata può a sua volta venir derivata termine a termine e ciò tante volte quante si vuole. Dunque:

Corollario 67 *La somma di una serie di potenze di raggio di convergenza non nullo è una funzione di classe C^∞ nell'intervallo di convergenza.*

Chiediamoci ora come sia possibile calcolare il raggio di convergenza di una serie di potenze. Esiste una formula per il raggio di convergenza, che non possiamo presentare. Possiamo però presentare due test particolari, che si ottengono per confronto con la serie geometrica. Il primo si può applicare quando si ha $a_n \neq 0$ per ogni n (e basta che questa condizione sia soddisfatta per n maggiore di un opportuno N_0).

Teorema 68 *Supponiamo che $a_n \neq 0$ per ogni n e che esista, finito o meno,*

$$L = \lim \frac{|a_{n+1}|}{|a_n|}\,.$$

Allora il raggio di convergenza della serie è

$$\begin{cases} 0 & \text{se} \quad L = +\infty \\ \frac{1}{L} & \text{se} \quad 0 < L < +\infty \\ +\infty & \text{se} \quad L = 0\,. \end{cases}$$

Dim. Limitiamoci a considerare il caso $0 < L < +\infty$. Applichiamo il criterio del rapporto per la convergenza della serie di numeri

$$\sum a_n[x-x_0]^n\,,$$

con x fissato. Il criterio del rapporto asserisce che condizione sufficiente di convergenza è che per n sufficientemente grande valga

$$\frac{|a_{n+1}[x-x_0]^{n+1}|}{|a_n[x-x_0]^n|} = \frac{|a_{n+1}|}{|a_n|}|x-x_0| < q < 1\,. \tag{2.7}$$

La serie diverge se vale la disuguaglianza opposta, con $q > 1$.

Proviamo che si ha convergenza se

$$|x-x_0| < \frac{1}{L} \qquad \text{ossia se} \qquad L|x-x_0| < 1 \quad \text{(disuguaglianze strette)}.$$

In tal caso, $\tilde{q} \in [0,1)$ tale che

$$L|x-x_0| = \tilde{q} < 1 \qquad \text{e quindi} \qquad \lim \frac{|a_{n+1}|}{|a_n|}|x-x_0| = L|x-x_0| = \tilde{q} < 1\,. \tag{2.8}$$

Sia $q \in (\tilde{q}, 1)$. Il teorema di permanenza del segno mostra l'esistenza di un numero N tale che per ogni $n > N$ vale

$$\frac{|a_{n+1}|}{|a_n|}|x-x_0| < q\,.$$

Essendo $q \in (0,1)$, la serie converge (si ricordi (2.7)).

In modo analogo si trattano gli altri casi. ∎

Ripetiamo che il teorema precedente non può usarsi se infiniti coefficienti a_n sono nulli.

Usando il criterio della radice invece del criterio del rapporto si prova invece:

Teorema 69 *Se esiste, finito o meno, il limite*

$$\lim \sqrt[n]{|a_n|} = L$$

allora il raggio di convergenza è

$$R = \begin{cases} 0 & \text{se} \quad L = +\infty \\ 1/L & \text{se} \quad 0 < L < +\infty \\ +\infty & \text{se} \quad L = 0\,. \end{cases}$$

Si noti che il Teorema 69 può usarsi anche se infiniti coefficienti a_n sono nulli. Mostriamo ora che esistono serie di potenze il cui raggio di convergenza non può determinarsi né col criterio del rapporto né con quello della radice[4]

Esempio 70 Si consideri la serie

$$\sum_{n=0}^{+\infty} 2^{(-1)^n n} x^n\,.$$

E'

$$\sqrt[n]{|a_n|} = \begin{cases} 2 & \text{se } n \text{ è pari} \\ \frac{1}{2} & \text{se } n \text{ è dispari.} \end{cases}$$

[4]nella forma enunciata sopra. Il criterio della radice può modificarsi in modo da avere un nuovo criterio, che identifica in raggio di convergenza in ogni caso.

Quindi,

$$\lim \sqrt[n]{|a_n|} \qquad \text{non esiste}\,.$$

E':

$$\frac{a_{n+1}}{a_n} = \frac{2^{(-1)^{n+1}(n+1)}}{2^{(-1)^n n}} = \begin{cases} 2^{-2n+1} & \text{se } n \text{ è pari} \\ 2^{2n-1} & \text{se } n \text{ è dispari} \end{cases}$$

e quindi

$$\lim \frac{a_{n+1}}{a_n} \qquad \text{non esiste.}$$

In casi di questo tipo, conviene considerare x come una parametro e studiare la convergenza della serie numerica ottenuta, per ogni valore di x. Usando il criterio della radice per ogni x fissato, si vede che

$$\sqrt[n]{|a_n x^n|} < q < 1$$

se $2x < q$ e quindi se $x < q/2$, per qualche $q < 1$. Dunque il raggio di convergenza è almeno $1/2$. D'altra parte, se $|x| > 1/2$ allora per $n = 2k$, pari, si trova

$$a_{2k}x^{2k} = (2x)^{2k} \to +\infty\,.$$

Quindi, il termine generale non tende a zero e la serie non converge. Quindi, il raggio di convergenza è uguale ad $1/2$. ∎

2.4.1 Operazioni sulle serie di potenze

Le operazioni sulle serie di potenze si eseguono con le regole delle operazioni sulle serie numeriche, per ogni fissato valore del parametro x. Ci interessa però notare esplicitamente:

Somma di serie Siano $\sum a_n(x-x_0)^n$, $\sum b_n(x-x_0)^n$ due serie di potenze col medesimo centro x_0 e con raggio di convergenza rispettivamente R_a ed R_b. Sia

$$R = \min\{R_a\,,\ R_b\}\,.$$

Allora, ambedue le serie convergono per $|x-x_0| < R$ e quindi la loro somma può scriversi nuovamente come serie di potenze di centro x_0, convergente per $|x-x_0| < R$:

$$\left[\sum a_n(x-x_0)^n\right] + \left[\sum a_n(x-x_0)^n\right] = \sum [a_n + b_n](x-x_0)^n\,.$$

Il raggio di convergenza della serie somma può essere maggiore di R. Per esempio, se $b_n = -a_n$, allora la serie somma ha raggio di convergenza $+\infty$ anche se $R < +\infty$. Ciò però può aversi solamente se $R_1 = R_2$. Infatti, se per esempio $R_1 < R_2$ allora $R = R_1$. Infatti, scrivendo

$$\sum a_n x_n = \left[\sum (a_n + b_n)x^n\right] + \left[\sum (-b_n)x^n\right]$$

si vede che il raggio di convergenza della prima serie è almeno uguale a

$$\min\{R, R_2\}$$

e questo sarebbe maggiore di R_1 se fosse $R > R_1$.

L'operazione di somma può usarsi anche "al contrario" per calcolare il raggio di convergenza in certi casi nei quali falliscono sia il criterio del rapporto che quello della radice:

Esempio 71 Torniamo a considerare la serie vista all'esempio 70. La serie di quest'esempio si può vedere come somma delle due serie

$$\left[\sum a_n x^n\right] + \left[\sum b_n x^n\right] , \qquad a_n = \begin{cases} 2^n & \text{n pari} \\ 0 & \text{altrimenti} \end{cases} \qquad b_n = \begin{cases} 0 & \text{n pari} \\ 2^{-n} & \text{altrimenti} . \end{cases}$$

Ora, rinominando $2k$ l'indice n della prima serie, si trova

$$\sum 2^{2k}[x^2]^k .$$

Sostituendo $t = x^2$ si trova la serie di potenza

$$\sum 2^{2k} t^k ,$$

a cui si può applicare sia il criterio del rapporto che quello della radice. Il raggio di convergenza è 4 per la serie della variabile t, e quindi 2 per la serie della variabile x. Analogamente, la seconda serie converge per $|x| < 1/2$. Dunque, il raggio di convergenza della serie somma è esattamente $1/2$. ∎

Unicità della serie di potenze Mostriamo che la serie di potenze che converge ad una funzione $f(x)$, se esiste, è unica:

Teorema 72 *Siano*

$$\sum_{n=0}^{+\infty} a_n(x-x_0)^n , \qquad \sum_{n=0}^{+\infty} b_n(x-x_0)^n$$

due serie di potenze con raggio di convergenza rispettivamente $R_1 > 0$ ed $R_2 > 0$. Sia $0 < R < \min\{R_1 ,\ R_2\}$. Se

$$\sum_{n=0}^{+\infty} a_n(x-x_0)^n = \sum_{n=0}^{+\infty} b_n(x-x_0)^n$$

per ogni x con $|x - x_0| < R$ allora $a_n = b_n$ per ogni n e quindi $R_1 = R_2$.

Dim. Basta provare che se

$$\sum_{n=0}^{+\infty}(a_n - b_n)(x-x_0)^n = 0$$

per $|x - x_0| < R$ allora $a_n - b_n = 0$. Posto $x = x_0$ si trova infatti $a_0 - b_0 = 0$ e quindi l'uguaglianza precedente si scrive

$$0 = \sum_{n=1}^{+\infty}(a_n - b_n)(x-x_0)^n = (x-x_0)\sum_{n=1}^{+\infty}(a_n - b_n)(x-x_0)^{n-1} .$$

Dunque si ha anche

$$\sum_{n=1}^{+\infty}(a_n - b_n)(x-x_0)^{n-1} = 0 .$$

Calcolando per $x = x_0$ si vede $a_1 - b_1 = 0$ e dunque

$$\sum_{n=2}^{+\infty}(a_n - b_n)(x-x_0)^{n-1} = 0 .$$

Il procedimento può quindi ripetersi trovando $a_2 - b_2 = 0$, e così via. ∎

Prodotto alla Cauchy di serie di Potenze Consideriamo le due serie di potenze col medesimo centro x_0

$$\sum_{n=0}^{+\infty} a_n(x-x_0)^n\,, \qquad \sum_{k=0}^{+\infty} b_k(x-x_0)^k\,.$$

Calcoliamo tutti i possibili prodotti di un termine della prima con un termine della seconda e raccogliamo i coefficienti dei termini con la medesima potenza $(x-x_0)^r$. Si trova $(x-x_0)^r$ quando si fa il prodotto di $b_k(x-x_0)^k$ con tutti i termini della forma

$$a_{r-k}(x-x_0)^{r-k}\,.$$

Ossia, il coefficiente di $(x-x_0)^r$ viene ad essere

$$\sum_{k=0}^{r} b_k a_{r-k}\,.$$

Ciò suggerisce la definizione di *prodotto alla Cauchy*, definito da

$$\left[\sum_{n=0}^{+\infty} a_n(x-x_0)^n\right]\left[\sum_{k=0}^{+\infty} b_k(x-x_0)^k\right] = \sum_{r=0}^{+\infty}\left[\sum_{k=0}^{r} b_k a_{r-k}\right](x-x_0)^r\,.$$

E' appena il caso di notare che il "nome" degli indici non ha effetto sulla somma della serie. Si sono usati "nomi" diversi (n, k ed r) solo per chiarezza.

Il teorema che si può provare è il seguente:

Teorema 73 *Due serie di potenze abbiamo raggio di convergenza rispettivamente r_1 ed r_2 coan $r_1 \le r_2$. La serie prodotto alla Cauchy ha raggio di convergenza almeno uguale ad r_1.*

Osservazione 74 Consideriamo l'esempio seguente:

$$\left[\sum_{n=0}^{+\infty} x^{2n}\right]\cdot\left[\sum_{k=0}^{+\infty} x^{5k}\right]$$

Il prodotto alla Cauchy contiene potenze x^r solo se r può scriversi nella forma

$$r = 2n + 5k$$

per certi valori di n e k. Il coefficiente di questa potenza è la somma di tanti numeri 1 quanti sono i modi di scrivere r in tale forma.

Ma ora,

$$\sum_{n=0}^{+\infty} x^{2n} = \frac{1}{1-x^2}\,, \qquad \sum_{k=0}^{+\infty} x^{5k} = \frac{1}{1-x^5}$$

e il loro prodotto è

$$\frac{1}{1-x^2-x^5+x^7}\,.$$

Vedremo, al paragrafo 2.4.3, un modo diverso per rappresentare questa funzione come somma di una serie di potenze, che calcola i coefficienti della (unica!) espansione in serie di potenze facendo derivate. precisamente vedremo che il coefficiente c_r di x^r è

$$c_r = \frac{1}{r!} D^r_{x_0=0} \left[\frac{1}{1 - x^2 - x^5 + x^7} \right] .$$

In questo modo, dato r, si calcola facilmente in quanti modi questo si può rappresentare come somma di tante "monete" di valore 2 e tante "monete" di valore 5. ∎

2.4.2 Serie di potenze nel campo complesso

Le operazioni algebriche e le operazioni di limite possono anche calcolarsi tra numeri complessi. Quindi, le serie di potenze possono trattarsi anche nel campo complesso. Il Teorema 61 e il Teorema 63 si riformulano come segue:

Teorema 75 (*Teorema di Abel*) *Una serie di potenze*

$$\sum a_n (z - z_0)^n ,$$

nel campo complesso, converge sempre nel punto z_0. Se converge anche in altri punti, allora essa converge in un disco aperto di centro z_0 e raggio $R > 0$ (si scrive $R = +\infty$ se la serie converge per ogni z).

La convergenza è uniforme in ogni disco di centro z_0 e raggio $R_0 < R$; è assoluta in ciascun punto z interno al disco di convergenza.

Per *convergenza assoluta* nel caso di serie di potenze a valori complessi, si intende la convergenza della serie[5]

$$\sum |a_n| \cdot |z - z_0|^n .$$

La dimostrazione è del tutto analoga a quella vista per le serie di potenze reali. Però, ora può mostrarsi di più: sia $f(z)$ la somma di una serie di potenze di raggio di convergenza R. Esiste almeno un punto z_1 tale che $|z_1 - z_0| = R$, nel quale la funzione $f(z)$ non è regolare. Ciò chiarisce la stranezza notata all'Esempio 64. Per $|x| < 1$ vale

$$\frac{1}{1 + x^2} = \sum_{n=0}^{+\infty} (-1)^n x^{2n}$$

e la funzione $1/(1+x^2)$ è di classe $C^\infty(\mathbf{R})$. Non si vede quindi quale fenomeno possa bloccare la convergenza della serie. Se però la serie si legge nel campo complesso, allora

$$\frac{1}{1 + z^2} = \sum_{n=0}^{+\infty} (-1)^n z^{2n}$$

e la funzione $1/(1 + z^2)$ non è definita nei punti $\pm i$, che distano 1 dal centro della serie.

---

[5] ora $|\cdot|$ indica il modulo dei numeri complessi.

2.4.3 Serie di Taylor

Sia $f(x)$ una funzione di classe C^∞ in un intorno di x_0. Ad essa può associarsi la *serie di Taylor*

$$\sum_{n=0}^{+\infty} \frac{1}{n!} f^{(n)}(x_0)[x - x_0]^n\,.$$

Questa si chiama la *serie di Taylor* della funzione f. Questa serie può non convergere e, se converge, può non convergere alla funzione f, come mostra l'esempio seguente:

Esempio 76 Sia

$$f(x) = \begin{cases} e^{-1/x^2} & \text{se} \quad x \neq 0 \\ 0 & \text{se} \quad x = 0\,. \end{cases}$$

Questa funzione è di classe C^∞ su $\mathbf{R}$ e le sue derivate in $x_0 = 0$ sono tutte nulle. Dunque la serie di Taylor di centro 0 associata ad $f(x)$ ha tutti i coefficienti nulli: converge su $\mathbf{R}$ alla funzione identicamente zero e non ad f. ∎

Ci possiamo chiedere quindi sotto quali condizioni la serie di Taylor di f effettivamente converga ad f. Scrivendo la formula di Taylor di $f(x)$ arrestata all'ordine k e col resto in forma di Lagrange, si vede che

$$f(x) = \sum_{n=0}^{k} \frac{1}{n!} f^{(n)}(x_0)[x - x_0]^n + \frac{1}{(k+1)!} f^{(k+1)}(s_k)[x - x_0]^{k+1}$$

dove s_k dipende da k ed è compreso tra x_0 ed x. La serie di Taylor converge ad f quando il resto converge a zero. Una condizione perché ciò accada è:

Teorema 77 *Esistano M, L tali che*

$$|f^{(k)}(x)| < ML^k \qquad \forall x \in [x_0 - r, x_0 + r]\,.$$

La serie di Taylor di $f(x)$ converge su $[x_0 - r, x_0 + r]$ e converge alla funzione $f(x)$.

Dim. Ricordiamo che per ogni x si ha

$$\lim_{n \to +\infty} \frac{x^n}{n!} = 0\,.$$

Si osservi ora che

$$\frac{1}{(k+1)!} \left| f^{(k+1)}(s_k)[x - x_0]^{k+1} \right| < M \frac{(Lr)^{k+1}}{(k+1)!}.$$

Il membro destro tende a zero e quindi, per il criterio del confronto, tende a zero uniformemente anche l'errore

$$\left| f(x) - \sum_{n=0}^{k} \frac{1}{n!} f^{(n)}(x_0)[x - x_0]^n \right| = \frac{1}{(k+1)!} \left| f^{(k+1)}(s_k)[x - x_0]^{k+1} \right|\,. \qquad ∎$$

La condizione del Teorema 77 è soddisfatta nel caso delle funzioni di cui correntemente si usano gli sviluppi di Taylor, almeno su un opportuno intervallo. La tabella seguente riporta alcune funzioni e il raggio di convergenza della relativa serie di McLaurin (ossia, della serie di Taylor di centro 0).

Funzione	Raggio di conv.	Serie
e^x	$+\infty$	$\sum_{n=0}^{\infty} \frac{x^n}{n!}$
$\sin x$	$+\infty$	$\sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$
$\cos x$	$+\infty$	$\sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!}$
$\sinh x$	$+\infty$	$\sum_{n=0}^{+\infty} \frac{x^{2n+1}}{(2n+1)!}$
$\cosh x$	$+\infty$	$\sum_{n=0}^{+\infty} \frac{x^{2n}}{(2n)!}$
$\log(1+x)$	1	$\sum_{n=1}^{+\infty} (-1)^{n+1} \frac{x^n}{n}$
$(1+x)^\alpha$	1	$\sum_{n=0}^{+\infty} \binom{\alpha}{n} x^n$

Infine, studiamo la relazione tra parità di una funzione e sua serie di Taylor. Vale:

Teorema 78 *Sia*

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n \,.$$

La funzione è pari se e solo se sono nulli i coefficienti a_n di indice dispari; la funzione è dispari se e solo se sono nulli i coefficienti a_n di indice pari.

Dim. Sia $f(x)$ sviluppabile in serie di Taylor. Se è dispari, essa si annulla in 0. Dunque $a_0 = 0$. La sua derivata prima è pari ma la derivata seconda è dispari e quindi $f''(0) = 0$ ossia vale anche $a_2 = 0$. Iterando questo procedimento si vede che ogni a_{2k} è nullo.

Se $f(x)$ è pari, sono dispari tutte le sue derivate di ordine dispari e quindi sono nulli gli a_{2k+1}.

Il viceversa è ovvio. ∎

2.4.4 Serie di potenze ed equazioni differenziali lineari

Consideriamo il problema di Cauchy

$$x' = ax \,, \qquad x(0) = x_0 \,.$$

Il coefficiente a è costante. Per definizione, la soluzione x è continua e quindi, dall'uguaglianza, è addiritture continuamente derivabile; e quindi

$$x'' = ax' = a^2 x \,.$$

Così proseguendo,

$$x^{(n)} = a^n x$$

e quindi, per $t = 0$,

$$x^{(n)}(0) = a^n x_0 \,.$$

Dunque, la soluzione $x(t)$ è di classe C^∞ e verifica le condizioni del Teorema 77 su tutti gli intervalli chiusi contenenti x_0. Dunque, la soluzione si esprime in forma di serie di potenze

$$x(t) = \sum_{n=0}^{+\infty} \frac{1}{n!} a^n t^n$$

D'altra parte si verifica immediatamente che questa è la serie dell'esponenziale e quindi si ritrova il risultato noto

$$x(t) = e^{at}x_0 \,.$$

Consideriamo ora il *sistema di equazioni differenziali lineari*

$$\mathbf{x}' = A\mathbf{x} \tag{2.9}$$

ove $\mathbf{x}$ è un vettore di $\mathbf{R}^n$ ed A è una matrice $n \times n$ che è costante, ossia non dipende né da t né da x. Vogliamo rappresentare la soluzione di questo sistema che verifica l'ulteriore condizione

$$\mathbf{x}(t_0) = \mathbf{x}_0 \,.$$

E' facile vedere che tutto ciò che abbiamo detto sulle serie di potenze e sulle serie di Taylor si estende senza cambiamenti a funzioni a valori vettori o matrici e quindi è ancora vero che

$$\mathbf{x}(t) = \left(\sum_{n=0}^{+\infty} \frac{1}{n!} A^n (t-t_0)^n \right) \mathbf{x}_0 \,.$$

Questa serie si indica col simbolo

$$e^{A(t-t_0)} = \sum_{n=0}^{+\infty} \frac{1}{n!} A^n (t-t_0)^n \,. \tag{2.10}$$

Ciò definisce l'*esponenziale di una matrice*, e la (2.10) rappresenta in forma compatta le soluzioni dell'equazione differenziale lineare a coefficienti costanti (2.9):

Osservazione 79 Va notato un fatto importante: l'esponenziale di matrice può essere un polinomio. Per esempio, se

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

allora $A^2 = 0$ e quindi

$$e^{At} = \begin{bmatrix} 1 & t \\ 1 & 0 \end{bmatrix} :$$

e^{At} è un polinomio di primo grado. Si prova invece che se

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

allora

$$e^{At} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} . \quad \blacksquare$$

Proprietà importanti della matrice esponenziale e^{At} sono espresse dal teorema seguente, che non proviamo:

Teorema 80 *Vale:*

- $Ae^{At} = e^{At}A$.

- $\det e^A = \exp\{\sum_{i=1}^n a_{ii}\}$. *Dunque,* $\det e^A$ *è sempre diverso da zero: la matrice* e^A *è invertibile per ogni* A.
- $\left[e^A\right]^{-1} = e^{-A}$.
- *Se* $AB = BA$ *allora* $e^A e^B = e^{A+B}$. *In particolare, vale sempre* $e^{At}e^{At'} = e^{A(t+t')}$.
- *La funzione* $t \to e^{At}$ *è derivabile e*

$$\frac{d}{dt}e^{At} = Ae^{At}\,.$$

L'introduzione dell'esponenziale e^{At} della matrice At permette anche di rappresentare la soluzione del problema

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}(t) \qquad \mathbf{x}(t_0) = \mathbf{x}_0\,.$$

Procediamo esattamente come già si è visto (nel corso di Analisi Matematica 1) per l'equazione scalare: moltiplicando i due membri per e^{-At} si trova

$$e^{-At}\mathbf{x}'(t) - Ae^{-At}\mathbf{x}(t) = e^{-At}\mathbf{f}(t)\,. \tag{2.11}$$

La regola della derivata del prodotto si estende al prodotto di una matrice per un vettore[6] e quindi la (2.11) è

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-At}\mathbf{x}(t)\right) = e^{-At}\mathbf{f}(t)\,.$$

Integrando i due membri da t_0 a t si trova

$$e^{-At}\mathbf{x}(t) - e^{-At_0}\mathbf{x}_0 = \int_{t_0}^t e^{-As}\mathbf{f}(s)\,\mathrm{d}s\,. \tag{2.12}$$

Moltiplichiamo i due membri di (2.12) per e^{At} e usiamo le proprietà nel teorema 80. Si trova

$$\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{x}_0 + \int_0^t e^{A(t-s)}\mathbf{f}(s)\,\mathrm{d}s\,.$$

2.5 Serie di Fourier: introduzione

Oltre alle serie di potenze, nelle applicazioni si incontrano molti altri tipo di "serie di funzioni", la cui teoria comunque è sostanzialmente più complessa e viene qui esaminata per sommi capi nel caso di gran lunga più importante delle serie di Fourier.

Si chiamano *serie di Fourier* le serie del tipo

$$a_0 + \sum_{n=1}^{+\infty}\left[a_n \cos n\frac{\pi}{L}x + b_n \sin n\frac{\pi}{L}x\right]\,. \tag{2.13}$$

I coefficienti a_n e b_n sono reali.

Se $L = \pi$ si trova

$$a_0 + \sum_{n=1}^{N}\left[a_n \cos nx + b_n \sin nx\right]\,.$$

---

[6] con l'avvertenza di non commutare i fattori!

Si noti che, usando $\sin 0x = 0$, si potrebbe assorbire il coefficiente a_0 nella serie scritta con $n \geq 0$ invece che con $n \geq 1$. Vedremo che c'è una buona ragione per separare a_0 dagli a_n con $n > 0$.

Ovviamente una serie di Fourier non sempre converge. La convergenza sarà implicata da opportune proprietà dei coefficienti a_n e b_n. Per esempio, certamente si ha convergenza (uniforme) quando $a_n = b_n = q^n$, con $|q| < 1$. Il problema della convergenza puntuale o uniforme delle serie di Fourier comunque è assai delicato e lo illustreremo più avanti. Per ora notiamo che se la serie converge per un $x_0 \in [-L, L]$ allora essa converge in ogni punto x_0 di forma $x_0 + 2kL$ con k intero; e se converge per ogni $x \in [-L, L]$ essa converge per ogni $x \in \mathbf{R}$ e converge ad una funzione peridica di periodo $T = 2L$. Per questa ragione, prima di studiare le serie di Fourier, vogliamo richiamare alcune proprietà delle funzioni periodiche.

2.5.1 Premesse: le funzioni periodiche

Sia $f(x)$ una funzione della variabile reale x. Si dice che $f(x)$ è periodica di periodo T quando:

- E' $T \neq 0$.
- la funzione $f(x)$ è definita in $x + T$ se e solo se è definita in x. E' conseguenza di questo che la funzione è definita in $x + nT$ (ove n è intero qualsiasi) se e solo se è definita in x.
- per ogni x nel dominio della funzione, si ha $f(x) = f(x + T)$ e quindi anche $f(x) = f(x + nT)$ per ogni numero intero n.

Teorema 81 *Sia $f(x)$ periodica di periodo T e sia S numero reale. La funzione $f(Sx)$ ha periodo T/S. In particolare, se $S = T/(2\pi)$, la funzione ha periodo 2π.*

La verifica è immediata:

$$f(S(x + T/S)) = f(T + Sx) = f(Sx) \,.$$

Una funzione periodica non ha un solo periodo: se T è un periodo anche $2T$, $-T$, $-2T$ ecc. sono periodi. L'insieme dei periodi positivi ha però un estremo inferiore che può essere nullo[7]. Molto spesso, quando si parla di "periodo" di una funzione periodica si intende che esista un minimo periodo positivo, e per *periodo* si intende tale numero[8]. Se T è il (minimo) periodo di $f(x)$, allora $1/T$ si chiama la *frequenza* di $f(x)$ mentre $2\pi/T$ si chiama la *frequenza angolare* di $f(x)$.

Per esempio, le funzioni $\sin \omega x$ e $\cos \omega x$ sono funzioni periodiche di periodo $2\pi/\omega$ ovunque definite mentre $\tan \omega x$ è una funzione di periodo π/ω, che però non è ovunque definita.

[7] L'estremo inferiore dei periodi può essere nullo. Accade ciò se la funzione è costante, ma non solo in tal caso. Per esempio, l'estremo inferiore dei periodi positivi è nullo per la funzione di Dirichlet.

[8] molto spesso, ma non sempre: si faccia attenzione al contesto!

Osservazione 82 Vediamo come i concetti precedenti si applicano alla serie di Fourier (2.13). Le funzioni $\cos n(\pi/L)x$ e $\sin n(\pi/L)x$ hanno (minimo) periodo $2L/n$ e quindi tutti i termini della serie di Fourier hanno periodo (non necessariamente minimo) $T = 2L$. Per questa ragione potremo limitarci a studiare la serie di Fourier sull'intervallo $[-L, L]$ o su un qualsiasi altro intervallo di lunghezza $2L$, per esempio $[0, 2L]$.

La frequenza di $\cos n(\pi/L)x$ e $\sin n(\pi/L)x$ è $n/2L$ mentre la loro frequenza angolare è

$$2\pi\frac{1}{T} = n\frac{\pi}{L} \ :$$

il coefficiente di x nei due termini in posizione n della serie di Fourier è la frequenza angolare $n\pi/L$. ∎

Vale:

Teorema 83 *Sia $f(t)$ continua su $\mathbf{R}$ e periodica di periodo T. Per ogni $x \in \mathbf{R}$ si ha*

$$\int_0^T f(s)\, ds = \int_x^{x+T} f(s)\, ds\,, \qquad \int_0^T f(x+s)\, ds = \int_0^T f(s)\, ds\,.$$

Dim. Conviene provare la prima uguaglianza in due passi:

Passo 1. Proviamo che per ogni k intero si ha:

$$\int_0^T f(s)\, \mathrm{d}s = \int_{kT}^{(k+1)T} f(s)\, \mathrm{d}s\,;$$

ossia proviamo l'uguaglianza richiesta nel caso particolare $x = kT$. Per questo sostituiamo $s = kT + r$ nell'integrale a destra. Si trova

$$\int_{kT}^{(k+1)T} f(s)\, \mathrm{d}s = \int_0^T f(kT+r)\, \mathrm{d}r = \int_0^T f(r)\, \mathrm{d}r\,.$$

L'ultima uguaglianza vale perché la funzione è periodica di periodo T e quindi $f(kT+r) = f(r)$.

Passo 2. Consideriamo ora il caso generale

$$kT < x < (k+1)T < x+T\,.$$

Proviamo che vale

$$\int_x^{x+T} f(s)\, \mathrm{d}s = \int_{kT}^{(k+1)T} f(s)\, \mathrm{d}s\,.$$

Ciò fatto, l'uguaglianza richiesta seguirà dal **passo 1.**

Per capire questa parte del calcolo, si guardi la figura 2.8. Come suggerito dalla figura, scriviamo

$$\int_x^{x+T} f(s)\, \mathrm{d}s = \int_x^{(k+1)T} f(s)\, \mathrm{d}s + \int_{(k+1)T}^{x+T} f(s)\, \mathrm{d}s\,. \tag{2.14}$$

La sostituzione $s = T + r$ mostra che

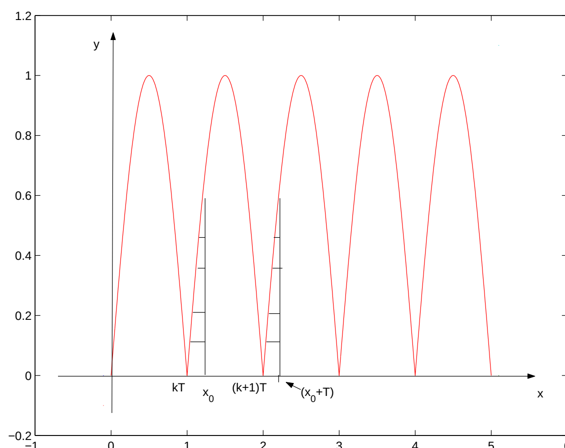
$$\int_{(k+1)T}^{x+T} f(s)\,\mathrm{d}s = \int_{kT}^{x} f(T+r)\,\mathrm{d}r = \int_{kT}^{x} f(r)\,\mathrm{d}r\,.$$

L'ultima uguaglianza vale perché $f(x)$ è periodica di periodo T e quindi $f(r) = f(T+r)$. Tornando alla (2.14) si vede che

$$\int_{x}^{x+T} f(s)\,\mathrm{d}s = \int_{x}^{(k+1)T} f(s)\,\mathrm{d}s + \int_{kT}^{x} f(r)\,\mathrm{d}r = \int_{kT}^{(k+1)T} f(r)\,\mathrm{d}r\,,$$

come volevamo provare.

Figura 2.8: Illustrazione del Teorema 83



La seconda uguaglianza è conseguenza della prima perché

$$\int_0^T f(x+s)\,\mathrm{d}s = \int_x^{x+T} f(r)\,\mathrm{d}r = \int_0^T f(r)\,\mathrm{d}r\,.$$

Si noti che nel teorema precedente T non è necessariamente il minimo periodo. ∎

2.5.2 Premesse: le formule d'Eulero

Ricordiamo le formule d'Eulero, incontrate nello studio delle equazioni differenziali lineari:

$$e^{ix} = \cos x + i \sin x$$

e quindi

$$e^{-ix} = \cos x - i \sin x\,.$$

Osservazione 84 Le formule d'Eulero mostrano che la funzione $x \to e^{ix}$ è periodica di periodo 2π. E' anche vero che l'estensione della funzione esponenziale al piano complesso

$$e^{x+iy} = e^{x}(\cos y + i \sin y)$$

è periodica di periodo $2\pi i$, ossia è periodica quando il suo argomento si muove parallelamente all'asse immaginario. ∎

Sommando e sottraendo membro a membro, si trovano le uguaglianze

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}\,, \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2}\,.$$

Anch'esse vanno sotto il nome di *formule d'Eulero*. Sostituendo queste espressioni in

$$a_0 + \sum_{n=1}^{N} [a_n \cos nx + b_n \sin nx]$$

si trova (si ricordi che $-i = 1/i$)

$$a_0 + \sum_{n=1}^{N} \frac{a_n - ib_n}{2} e^{inx} + \sum_{n=1}^{N} \frac{a_n + ib_n}{2} e^{-inx} = \sum_{n=-N}^{N} c_n e^{inx}$$

ove ora i c_n sono i numeri complessi

$$\begin{cases} c_0 = a_0 \\ c_n = \frac{a_n - ib_n}{2} & \text{se} \quad n > 0 \\ c_n = \frac{a_n + ib_n}{2} & \text{se} \quad n < 0 \end{cases} \tag{2.15}$$

e quindi tali che

$$c_{-n} = \bar{c}_n\,.$$

Si osservi che anche in questa scrittura il termine con $n = 0$ ha un ruolo particolare: $c_0 = \bar{c}_0$ è reale.

Sostituendo in

$$a_0 + \sum_{n=1}^{N} \left[a_n \cos n\frac{\pi}{L}x + b_n \sin n\frac{\pi}{L}x \right] \tag{2.16}$$

si trova

$$\sum_{n=-N}^{N} c_n e^{in(\pi/L)x} \tag{2.17}$$

con i coefficienti c_n ancora dati da (2.15) e quindi con

$$c_{-n} = \bar{c}_n\,, \qquad \text{e} \qquad c_0 = \bar{c}_0\,. \tag{2.18}$$

E' facile vedere che si fa anche il passaggio inverso: se valgono le (2.18) allora nella (2.17) i termini puramente immaginari si elidono e la somma (2.17) si riduce a (2.16), con coefficienti reali.

Un'espressione della forma (2.17) si chiama un *polinomio trigonometrico*. Se valgono le (2.18), il polinomio trigonometrico (2.17) si può scrivere nella forma "reale" (2.16).

L'espressione (2.16) è una somma parziale della serie di Fourier (2.13). Per questa ragione, la serie di Fourier (2.13) si può anche scrivere in "forma complessa",

$$\sum_{n-=\infty}^{+\infty} c_n e^{in(\pi/L)x} . \tag{2.19}$$

Il fatto importante da ricordare è che se vogliamo che questa serie corrisponda alla (2.13) le somme parziali vanno prese in modo simmetrico: le somma parziali sono

$$\sum_{n=-N}^{N} c_n e^{in(\pi/L)x} \tag{2.20}$$

e non

$$\sum_{n=-K}^{N} c_n e^{in(\pi/L)x} \tag{2.21}$$

con K ed N tra loro indipendenti. E infatti può accadere che per $K \to -\infty$ ed $N \to +\infty$, indipendentemente, la (2.21) non ammetta limite nemmeno se (2.20) ammette limite per $N \to +\infty$.

Avremo bisogno di calcolare derivate e integrali di funzioni

$$f(x) + ig(x)$$

della variabile reale x, a valori numeri complessi. Per definizione,

$$\frac{\mathrm{d}}{\mathrm{d}x}[f(x) + ig(x)] = f'(x) + ig'(x) , \qquad \int_a^b [f(x) + ig(x)]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + i \int_a^b g(x)\,\mathrm{d}x$$

e quindi

$$\int_a^b \frac{\mathrm{d}}{\mathrm{d}x}[f(x) + ig(x)]\,\mathrm{d}x = [f(b) + ig(b)] - [f(a) + ig(a)] .$$

Essendo

$$\frac{\mathrm{d}}{\mathrm{d}x} e^{inx} = in e^{inx} ,$$

si trova:

$$\begin{cases} \displaystyle\int_{-\pi}^{\pi} \cos nx \cos mx\,\mathrm{d}x & = \begin{cases} 2\pi & \text{se} \quad n = m = 0 \\ 0 & \text{se} \quad n \neq m \\ \pi & \text{se} \quad n = m \neq 0 \end{cases} \\ \displaystyle\int_{-\pi}^{\pi} \sin nx \cos mx\,\mathrm{d}x & = 0 \text{ per ogni } n,\, m. \\ \displaystyle\int_{-\pi}^{\pi} \sin nx \sin mx\,\mathrm{d}x & = \begin{cases} 0 & \text{se} \quad n \neq m \\ \pi & \text{se} \quad n = m \end{cases} \end{cases} \tag{2.22}$$

Per verificare la prima delle uguaglianze precedenti (con $n \neq m$) calcoliamo, usando le formule d'Eulero

$$\int_{-\pi}^{\pi} \cos nx \cos mx\,\mathrm{d}x = \frac{1}{4} \int_{-\pi}^{\pi} \left[e^{inx} + e^{-inx}\right] \left[e^{imx} + e^{-imx}\right]\,\mathrm{d}x$$

$$= \frac{1}{4} \int_{-\pi}^{\pi} \left[e^{i(n+m)x} + e^{i(n-m)x} + e^{-i(n+m)x} + e^{-i(n-m)x}\right]\,\mathrm{d}x .$$

L'asserto ora segue perchè, essendo per esempio

$$\frac{\mathrm{d}}{\mathrm{d}x}e^{i(n+m)x} = i(n+m)e^{i(n+m)x}\,,$$

si ha

$$\int_{-\pi}^{\pi} e^{i(n+m)x}\,\mathrm{d}x = \frac{1}{i(n+m)}\left[e^{i(n+m)\pi} - e^{-i(n+m)\pi}\right] = \frac{2}{n+m}\sin((n+m)\pi) = 0\,.$$

Le altre uguaglianze si provano in modo analogo.

Osservazione 85 a queste conclusioni si può arrivare in modo più elementare se si ricordano le formule di Werner. il primo integrale si può calcolare ricordando che

$$[\cos nx][\cos mx] = \frac{1}{2}\left[\cos(n+m)x + \cos(n-m)x\right]\,. \quad \blacksquare$$

Più in generale, su un intervallo $[-L,L]$ vale

$$\left\{\begin{array}{lcl} \displaystyle\int_{-L}^{L} e^{in(\pi/L)x}\,\mathrm{d}x & = & \left\{\begin{array}{lll} 2L & \text{se} & n=0 \\ 0 & \text{se} & n\neq 0 \end{array}\right. \\ \displaystyle\int_{-L}^{L} \cos nx\cos mx\,\mathrm{d}x & = & \left\{\begin{array}{lll} 2L & \text{se} & n=m=0 \\ 0 & \text{se} & n\neq m \\ L & \text{se} & n=m \end{array}\right. \\ \displaystyle\int_{-L}^{L} \sin nx\cos mx\,\mathrm{d}x & = & 0 \text{ per ogni } n,\, m. \\ \displaystyle\int_{-L}^{L} \sin nx\sin mx\,\mathrm{d}x & = & \left\{\begin{array}{lll} 0 & \text{se} & n\neq m \\ L & \text{se} & n=m \end{array}\right. \end{array}\right. \tag{2.23}$$

Le uguaglianze precedenti mostrano che le funzioni

$$\frac{1}{\sqrt{2L}}, \qquad \frac{1}{\sqrt{L}}\cos n\frac{\pi}{L}x\,, \qquad \frac{1}{\sqrt{L}}\sin n\frac{\pi}{L}x\,,$$

equivalentemente

$$\frac{1}{\sqrt{2L}}e^{in(\pi/L)x}\,,$$

sono due a due ortogonali in $\mathcal{L}^2(-L,L)$ e tutte di norma uguale ad 1. Si dice brevemente che sono *sistemi ortonormali* in $\mathcal{L}^2(-L,L)$.

Quest'osservazione suggerisce che l'ambiente in cui è più facile studiare la serie di Fourier sia lo spazio $\mathcal{L}^2(-L,L)$ e non lo spazio $C(-L,L)$.

2.6 La serie di Fourier in $\mathcal{L}^2(-L,L)$

Non è stato possibile introdurre in modo rigoroso lo spazio $\mathcal{L}^2(-L,L)$ e ciò indica che lo studio della serie di Fourier è molto più complesso di quello delle serie di potenze, e può essere solo accennato.

Consideriamo il polinomio trigonometrico

$$\sum_{n=-N}^{N} c_n e^{in(\pi/L)x}\,, \qquad c_n = \bar{c}_{-n} \tag{2.24}$$

equivalentemente, se $c_n = a_n + ib_n = \bar{c}_{-n}$,

$$a_0 + \sum_{n=1}^{N} \left[a_n \cos n(\pi/L)x + b_n \sin n(\pi/L)x\right]\,. \tag{2.25}$$

Indichiamo questo polinomio trigonometrico col simbolo $P(x)$. Ovviamente, $P(x)$ è una funzione continua e periodica su $\mathbf{R}$. Se i valori di $P(x)$ su $[-L, L]$ si conoscono, da questi si ricavano facilmente i coefficienti c_n ed i coefficienti a_n e b_n. Infatti, moltiplicando i due membri dell'uguaglianza

$$P(x) = \sum_{n=-N}^{N} c_n e^{in(\pi/L)x}$$

per $e^{-ir(\pi/L)x}$ e integrando su $[-L, L]$ si trova 0 se $r \neq N$. Altrimenti si trova

$$c_r = \frac{1}{2L}\int_{-\pi}^{\pi} P(x)e^{-ir(\pi/L)x}\,\mathrm{d}x\,.$$

Analogamente,

$$\begin{cases} a_0 &= \frac{1}{2L}\int_{-L}^{L} P(x)\,\mathrm{d}x \\ a_k &= \frac{1}{L}\int_{-L}^{L} P(x)\cos k(\pi/L)x\,\mathrm{d}x \qquad (\text{se } k > 0) \\ b_k &= \frac{1}{L}\int_{-L}^{L} P(x)\sin k(\pi/L)x\,\mathrm{d}x\,. \end{cases}$$

(ossia, la formula per a_0 non si ottiene da quella di a_k ponendo $k = 0$. Per questa ragione conviene scrivere a_0 separato dalla sommatoria).

Vale inoltre:

Teorema 86 *E':*

$$\begin{cases} \dfrac{1}{2L}\displaystyle\int_{-L}^{L} |P(x)|^2\, dx = \sum_{n=-N}^{+N} |c_n|^2 \\ \dfrac{1}{2L}\displaystyle\int_{-L}^{L} |P(x)|^2\, dx = a_0^2 + \frac{1}{2}\sum_{n=1}^{+\infty}\left(a_n^2 + b_n^2\right)\,. \end{cases} \tag{2.26}$$

L'identità precedente va sotto il nome di *Identità di Parseval*.

Si noti il significato di a_0 e di c_0: questi numeri sono le medie integrali di $P(x)$. Analogamente, il membro sinistro dell'identità di Parseval è la media integrale di $|P(x)|^2$.

Sia ora $f(x) \in \mathcal{L}[-L, L]$. Chiamiamo *coefficienti di Fourier* della funzione $f(x)$ i numeri

$$c_r = \frac{1}{2L}\int_{-L}^{L} f(x)e^{-ir(\pi/L)x}\,\mathrm{d}x \tag{2.27}$$

(se vogliamo scrivere la serie di Fourier con gli esponenziali complessi),

$$\begin{cases} a_0 &= \frac{1}{2L}\int_{-L}^{L} f(x)\,\mathrm{d}x \\ a_k &= \frac{1}{L}\int_{-L}^{L} f(x)\cos k(\pi/L)x\,\mathrm{d}x \quad (\text{se } k>0) \\ b_k &= \frac{1}{L}\int_{-L}^{L} f(x)\sin k(\pi/L)x\,\mathrm{d}x \end{cases} \tag{2.28}$$

(se vogliamo scrivere la serie di Fourier nel campo reale).

Si chiama *serie di Fourier* associata alla funzione $f(x)$ la serie (2.13) con i coefficienti dati da (2.28) o, equivalentemente, la serie (2.19) con i coefficienti dati da (2.27).

Per fissare le idee e scrivere formule più semplici, usiamo ora la serie di Fourier scritta mediante gli esponenziali complessi. Considerazioni del tutto analoghe valgono anche per la serie di Fourier scritta nel campo reale.

Consideriamo la serie

$$\sum_{-\infty}^{+\infty} c_n e^{inx}$$

e la sua somma parziale N–ma

$$S_N(x) = \sum_{-N}^{+N} c_n e^{inx}\,.$$

Si noti che $S_N(x)$ è il polinomio trigonometrico (2.24) e che, se $c_n = a_n + ib_n = \bar{c}_{-n}$, allora $S_N(x)$ può scriversi nella forma (2.25).

Si può provare:

Teorema 87 *Sia $f \in \mathcal{L}^2(-L, L)$. Vale:*

$$\lim_{N\to+\infty} \int_{-L}^{L} |f(x) - S_N(x)|^2 \, dx = 0 \qquad \text{ossia} \qquad \lim_{N\to+\infty} ||f - S_N||_{\mathcal{L}^2(-L,L)} = 0\,.$$

Inoltre, la serie converge incondizionatamente, *ossia la serie converge, e sempre alla medesima funzione $f(x)$, anche riordinandone i termini.*

Dunque, la successione delle somme parziali $(S_N(x))$ converge ad $f(x)$ nella distanza di $\mathcal{L}^2(-L, L)$. Sottolineiamo nuovamente che il teorema riguarda $S_N(x)$ e non per esempio una somma $\sum_{n=-K}^{n=N} c_n e^{in(\pi/L)x}$. Anche se i c_n sono i coefficienti di Fourier di f, niente può dirsi del comportamento di questa serie per $N \to +\infty$, $K \to +\infty$ in modo indipendente.

Diamo un'interpretazione geometrica di $S_N(x)$. Consideriamo il sottospazio lineare $\mathcal{V}_N$,

$$\mathcal{V}_N = \left\{ \sum_{n=-N}^{+N} \gamma_n e^{in(\pi/L)x}\,, \quad \gamma_n = \bar{\gamma}_{-n}\,, \qquad \gamma_n \in \mathbf{C} \right\}$$

Che è uno spazio vettoriale di dimensione $2N + 1$. Si ha:

Teorema 88 *La somma parziale $S_N(x)$ di $f(x)$ è l'elemento di $\mathcal{V}_N$ che ha minor distanza da $f(x)$ nel senso della distanza di $\mathcal{L}^2(-L, L)$.*

Dim. Facciamo la dimostrazione nel caso $N = 1$. La dimostrazione nel caso generale è analoga.

Gli elementi dello spazio $\mathcal{V}_1$ sono le funzioni

$$\gamma_0 + \gamma_1 e^{ix} + \bar{\gamma}_1 e^{-ix} \qquad \gamma_i \in \mathbf{C}\,,$$

equivalentemente

$$\alpha_0 + \alpha_1 \cos(\pi/L)x + \beta_1 \sin(\pi/L)x\,, \qquad \alpha_0\,,\ \alpha_1\,,\ \beta_1 \in \mathbf{R}\,.$$

Tra queste funzioni dobbiamo trovare quella che ha minima distanza da $f(x)$. Si tratta quindi di studiare un problema di minimo al variare dei parametri complessi γ_0 e γ_1 o, equivalentemente, al variare dei parametri reali α_0, α_1, β_1. Dato che i problemi di minimo che si sono studiati sono quelli di funzioni di variabile reale, conviene studiare il minimo della funzione

$$\Phi(\alpha_0, \alpha_1, \beta_1) = \int_{-L}^{L} [f(x) - \alpha_0 - \alpha_1 \cos(\pi/L)x - \beta_1 \sin(\pi/L)x]^2 \,\mathrm{d}x\,.$$

Il minimo esiste, come conseguenza del Teorema di Weierstrass, perché la funzione

$$(\alpha_0, \alpha_1, \beta_1) \longrightarrow \Phi(\alpha_0, \alpha_1, \beta_1)$$

è continua e tende a $+\infty$ per $||(\alpha_0, \alpha_1, \beta_1)|| \to +\infty$. Per trovarlo, annulliamo le derivate prime[9]. Si trovano le condizioni

$$\begin{aligned}
&\int_{-L}^{L} [f(x) - \alpha_0 - \alpha_1 \cos(\pi/L)x - \beta_1 \sin(\pi/L)x] \,\mathrm{d}x = 0\\
&\int_{-L}^{L} [f(x) - \alpha_0 - \alpha_1 \cos(\pi/L)x - \beta_1 \sin(\pi/L)x] \cos(\pi/L)x \,\mathrm{d}x = 0\\
&\int_{-L}^{L} [f(x) - \alpha_0 - \alpha_1 \cos(\pi/L)x - \beta_1 \sin(\pi/L)x] \sin(\pi/L)x \,\mathrm{d}x = 0\,.
\end{aligned}$$

Usando le uguaglianze (2.22), si trova che le tre derivate parziali si annullano solamente quando

$$\begin{aligned}
\alpha_0 &= (1/2L) \int_{-L}^{L} f(x) \,\mathrm{d}x\\
\alpha_1 &= (1/L) \int_{-L}^{L} f(x) \cos x \,\mathrm{d}x\\
\beta_1 &= (1/L) \int_{-L}^{L} f(x) \sin x \,\mathrm{d}x\,;
\end{aligned}$$

ossia, il punto di $\mathcal{V}_1$ che meno dista da $f(x)$ è $S_1(x)$. ∎

[9] si può provare che è lecito derivare sotto il segno di integrale. Ciò è provato al paragrafo 141. Alternativamente, si può sviluppare il quadrato portare i coefficienti α_0, α_1 e β_1 fuori dagli integrali. Quindi si calcolano le derivate.

Dunque l'interpretazione della serie di Fourier in $\mathcal{L}^2(-L, L)$ è la seguente: per ogni N si considera il sottospazio $\mathcal{V}_N$ di dimensione finita $2N + 1$ di $\mathcal{L}^2(-L, L)$. Si scrive la serie di Fourier di $f(x)$ e si tronca all'indice N. Si trova un elemento di $\mathcal{V}_N$ che è proprio l'elemento che meglio approssima la funzione $f(x)$ nel senso di $\mathcal{L}^2(-L, L)$. Usando una terminologia della geometria elementare, diremo che $S_N(x)$ è la *proiezione ortogonale* di $f(x)$ su $\mathcal{V}_N$. Il Teorema 87 si può riassumere dicendo che la successione delle proiezioni di f sui $\mathcal{V}_N$ converge ad f in $\mathcal{L}^2(-L, L)$.

Diciamo infine che *l'identità di Parseval* vale per ogni $f(x) \in \mathcal{L}^2(-L, L)$:

$$\frac{1}{2L}\int_{-L}^{L} |f(x)|^2 \,\mathrm{d}x = \sum_{n=-\infty}^{+\infty} |c_n|^2$$

ossia

$$\frac{1}{2L}\int_{-L}^{L} |f(x)|^2 \,\mathrm{d}x = a_0^2 + \frac{1}{2}\sum_{n=1}^{+\infty}[a_n^2 + b_n^2]\,.$$

Di conseguenza:

Teorema 89 *La successione dei coefficienti di Fourier tende a zero.*

L'identità di Parseval ha un'interpretazione importante per le applicazioni, che illustriamo con riferimento alla forma complessa, che scriviamo come:

$$\int_{-L}^{L} |f(x)|^2 \,\mathrm{d}x = 2L\left[\sum_{n=-\infty}^{+\infty} |c_n|^2\right]\,.$$

Interpretiamo la variabile x come "posizione" ed $f(x)$ come "velocità". Il primo integrale si interpreta come "energia" per esempio cinetica: la "somma" delle energie associate ad ogni particella del corpo.

La "componente" di frequenza $n/2L$, ossia

$$c_n e^{in(\pi/L)x}$$

ha quindi "energia" $2L|c_n|^2$. Quindi,

> l'energia totale ottenuta sommando le energie in tutte le posizioni è uguale alla somma delle energie delle componenti di tutte le frequenze.

Naturalmente, niente vieta che nella rappresentazione di un segnale $f(x)$ la componente di frequenza $n_0/2L$ abbia "energia" nulla, ossia che $c_{n_0} = 0$. Le considerazioni precedenti mostrano che l'energia di $f(x)$ si ripartisce tra i segnali e^{inx} per cui $c_n \neq 0$.

La successione $(n/2L, c_n)$ si chiama lo *spettro* del segnale.

Infine, notiamo che l'identità di Parseval mostra che se i coefficienti di Fourier sono tutti nulli allora la funzione è nulla, ed ovviamente vale anche il viceversa. Ossia:

Teorema 90 *Due funzioni f, g in $\mathcal{L}^2(-L, L)$ con i medesimi coefficienti di Fourier verificano*

$$\int_{-L}^{L} |f(x) - g(x)|^2 \; dx = 0\,. \tag{2.29}$$

Se accade che le due funzioni $f(x)$ e $g(x)$ sono continue, allora la condizione (2.29) implica che esse coincidono. In generale, se vale (2.29), le due funzioni possono "considerarsi equivalenti" nel senso che danno il medesimo contributo alle espressioni integrali nelle quali figurano.

Il calcolo dei coefficienti di Fourier

Il calcolo dei coefficienti di Fourier richiede il calcolo di integrali piuttosto noiosi. In certi casi si possono usare dei trucchi che semplificano il calcolo. Per esempio:

- Il prodotto di una funzione pari e di una dispari è dispari; e una funzione dispari ha integrale nullo su $[-L, L]$. Dunque, se $f(x)$ è dispari, i coefficienti a_n sono tutti nulli.

  Analogamente, se $f(x)$ è pari i coefficienti b_n sono tutti nulli.

- Si voglia calcolare la serie di Fourier della funzione $f(x) = x$. Questa è una funzione dispari e quindi vanno calcolati i soli coefficienti b_n,

  $$b_n = \frac{1}{L}\int_{-L}^{L} x \sin n\frac{\pi}{L}x \,\mathrm{d}x\,.$$

  Questi integrali si calcolano facilmente per parti ma ciò richiede un calcolo indipendente per ogni valore di n. Un modo più veloce è il seguente: si consideri la funzione

  $$s \longrightarrow x\sin sx \qquad \text{ossia} \qquad s \longrightarrow -\frac{\mathrm{d}}{\mathrm{d}s}\cos sx\,.$$

  Il numero Lb_n è il valore per $s = n\pi/L$ della funzione[10]

  $$-\int_{-L}^{L}\frac{\mathrm{d}}{\mathrm{d}s}\cos sx\,\mathrm{d}x = -\frac{\mathrm{d}}{\mathrm{d}s}\int_{-L}^{L}\cos sx\,\mathrm{d}x =$$
  $$-\frac{\mathrm{d}}{\mathrm{d}s}\left\{\frac{2}{s}\sin sL\right\} = \frac{2}{s^2}\sin sL - \frac{2L}{s}\cos sL$$

  (formula valida per $s \neq 0$). Calcolando per $s = n\pi/L$ e dividendo per L, si trova

  $$b_n = 2(-1)^{n+1}\frac{L}{n\pi}\,.$$

  Dunque,

  $$x = 2\frac{L}{\pi}\sum_{n=1}^{+\infty}\frac{(-1)^{n+1}}{n}\sin n\frac{\pi}{L}x\,. \tag{2.30}$$

  Procedendo in modo analogo si verifichi che

  $$x^2 = \frac{L^2}{3} - \frac{4L^2}{\pi^2}\left[\sum_{n=1}^{+\infty}\frac{(-1)^n}{n^2}\cos n\frac{\pi}{L}x\right]\,.$$

- Generalmente non è lecito derivare termine a termine una serie di Fourier. Invece, se $f(x) \in \mathcal{L}^2(-L, L)$, la serie di Fourier di $f(x)$ può integrarsi termine a termine: se

  $$f(x) = a_0 + \sum_{n=1}^{+\infty}\left[a_n \cos n(\pi/L)x + b_n \sin n(\pi/L)x\right]$$

[10] Lo scambio del segno di derivata con quello di integrale è giustificato al paragrafo 141.

allora

$$F(x)=\int_0^x f(s)\,\mathrm{d}s=a_0x+\frac{L}{\pi}\sum_{n=1}^{+\infty}\frac{b_n}{n}$$
$$+\frac{L}{\pi}\sum_{n=1}^{+\infty}\left[\frac{a_n}{n}\sin n(\pi/L)x-\frac{b_n}{n}\cos n(\pi/L)x\right].$$

Si noti che quella ottenuta non è una serie di Fourier, a causa del primo addendo a_0x. Se vogliamo la serie di Fourier di $F(x)$ dobbiamo esprimere anche la funzione x mediante la sua serie di Fourier, usando (2.30). In questo modo si trova

$$F(x)=\frac{L}{\pi}\sum_{n=1}^{+\infty}\frac{b_n}{n}$$
$$+\frac{L}{\pi}\sum_{n=1}^{+\infty}\left[\frac{2(-1)^{n+1}a_0+a_n}{n}\sin n(\pi/L)x-\frac{b_n}{n}\cos n(\pi/L)x\right].$$

2.6.1 Estensioni pari e dispari, e serie di Fourier

Il fatto seguente si è già notato: la funzione $f(x)$ sia pari,

$$f(x)=f(-x).$$

In questo caso, ciascuna delle funzioni

$$f(x)\sin nx$$

è dispari e quindi ha integrale nullo: i coefficienti b_n sono tutti nulli. Ovviamente, vale anche il viceversa se i coefficienti b_n sono tutti nulli, la somma della serie è una funzione pari. Analogamente, se $f(x)$ è dispari,

$$f(x)=-f(-x),$$

sono nulli i coefficienti a_n e viceversa. Dunque:

Teorema 91 *Sia*

$$f(x)=a_0+\sum_{n=1}^{+\infty}\left[a_n\cos n\frac{\pi}{L}x+b_n\sin n\frac{\pi}{L}x\right].$$

La funzione $f(x)$, periodica di periodo $T=2L$, è pari se e solo se $b_n=0$ per ogni n; è dispari se e solo $a_n=0$ per ogni n.

Supponiamo ora che sia data una funzione $f(x)$ definita solamente su $[0,L]$. Ad essa possono associarsi infinite serie di Fourier, una per ciascuna "arbitraria" estensione di $f(x)$ a $[-L,0]$. Tra queste estensioni due sono privilegiate: l'estensione pari e l'estensione dispari. L'estensione pari conduce ad una serie di soli coseni mentre l'estensione dispari conduce ad una serie di soli seni. Dunque:

Teorema 92 *Sia $f(x) \in \mathcal{L}^2(0,L)$. Essa è somma di un'*unica *serie di Fourier di soli seni, e di un'*unica *serie di Fourier di soli coseni. La prima definisce l'unica estensione dispari di $f(x)$ ad $\mathbf{R}$, periodica di periodo $2L$; la seconda definisce l'unica estensione pari di $f(x)$ ad $\mathbf{R}$, periodica di periodo $2L$.*

Le formule per la serie di Fourier di $f(x)$ di soli seni e di soli coseni si ottengono scrivendo nelle (2.28) le opportune estensioni di $f(x)$. Per esempio, usando l'estensione pari, per $k \neq 0$ si trova

$$a_k = (1/L)\int_{-L}^{L} f(x)\cos k(\pi/L)x \, \mathrm{d}x = \frac{2}{L}\int_0^L f(x)\cos k(\pi/L)x \, \mathrm{d}x \, .$$

Dato che queste formule (e la corrispondente identità di Parsevale) sono di uso molto comune, vanno ricordate. E' bene quindi scriverle in modo esplicito:

Sviluppo di Fourier in soli coseni

Sia $f(x) \in \mathcal{L}^2(0,L)$. Si ha:

$$f(x) = \alpha_0 + \sum_{n=1}^{+\infty} \alpha_n \cos n(\pi/L)x$$

con

$$\begin{cases} \alpha_0 &= (1/L)\int_0^L f(x)\,\mathrm{d}x \\ \alpha_n &= (2/L)\int_0^L f(x)\cos n(\pi/L)x\,\mathrm{d}x \quad (\text{se } n > 0) \end{cases} \tag{2.31}$$

La corrispondente identità di Parseval è

$$\frac{1}{L}\int_0^L |f(x)|^2 \, \mathrm{d}x = \alpha_0^2 + \sum_{n=1}^{+\infty} \alpha_n^2 = \sum_{n=0}^{+\infty} \alpha_n^2 \, .$$

Sviluppo di Fourier in soli seni

Sia $f(x) \in \mathcal{L}^2(0,L)$. Si ha:

$$f(x) = \sum_{n=1}^{+\infty} \beta_n \sin n(\pi/L)x$$

con

$$\beta_n = (2/L)\int_0^L f(x)\sin n(\pi/L)x \, \mathrm{d}x \, . \tag{2.32}$$

La corrispondente identità di Parseval è:

$$\frac{1}{L}\int_0^L |f(x)|^2 \, \mathrm{d}x = \sum_{n=1}^{+\infty} \beta_n^2 \, .$$

Osservazione 93 Sia $f(x) = \cos(\pi/L)x$, definita su $[0, L]$. Vogliamo calcolarne lo sviluppo di Fourier in soli coseni ed in soli seni. Lo sviluppo di Fourier in soli coseni è

$$f(x) = \cos(\pi/L)x\,.$$

Invece, per lo sviluppo di Fourier di soli seni dobbiamo calcolare i coefficienti b_n usando la formula (2.31). Ricordiamo la formula di Werner seguente:

$$(\sin a)(\cos b) = \frac{1}{2}\left[\sin(a+b) + \sin(a-b)\right]\,.$$

Si trova:

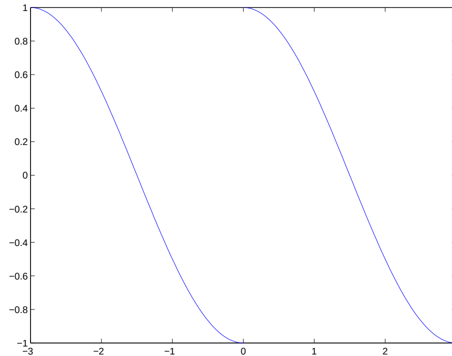
$$\begin{aligned}
\beta_n &= \frac{2}{L}\int_0^L \cos(\pi/L)x \sin n(\pi/L)x\,\mathrm{d}x \\
&= \frac{1}{L}\int_0^L \left\{\sin\frac{\pi}{L}(n+1)x + \sin\frac{\pi}{L}(n-1)x\right\}\,\mathrm{d}x \\
&= -\frac{1}{\pi(n+1)}\left\{\cos\pi(n+1) - 1\right\} - \frac{1}{\pi(n-1)}\left\{\cos\pi(n-1) - 1\right\} \\
&= \frac{4}{\pi}\begin{cases} \frac{n}{n^2-1} & \text{se } n \text{ è pari} \\ 0 & \text{se } n \text{ è dispari.} \end{cases}
\end{aligned}$$

Dunque,

$$\cos\frac{\pi}{L}x = \frac{8}{\pi}\sum_{k=1}^{+\infty}\frac{k}{4k^2-1}\sin 2k\frac{\pi}{L}x\,, \qquad 0 < x < L\,.$$

E' questa la serie di Fourier dell'estensione per periodicità della funzione $f(x)$ seguente, il cui grafico (con $L = 3$) è riportato in figura 2.9.

Figura 2.9: Estensione per periodicità della funzione (2.33)



$$f(x) = \begin{cases} \cos\frac{\pi}{L}x & \text{se} \quad 0 < x < L \\ -\cos\frac{\pi}{L}x & \text{se} \quad -L < x < 0 \end{cases} \quad \blacksquare \tag{2.33}$$

2.7 La convergenza puntuale della serie di Fourier

E' un fatto che la convergenza nel senso della norma di $\mathcal{L}^2(-L,L)$ non implica la convergenza puntuale, nemmeno in un solo punto. Anzi, si prova che esistono funzioni continue e periodiche su $[-L,L]$ la cui serie di Fourier non converge. Esistono però anche casi in cui la serie di Fourier converge puntualmente. Come abbiamo detto questo accade se, per esempio, $a_n = b_n = q^n$ con $|q| < 1$. Ci si può chiedere se sia possibile dare condizioni sulla funzione $f(x)$ che implichino la convergenza puntuale della serie di Fourier. Condizioni per questo sono note. In particolare si ha:

Teorema 94 *Sia $(a,b) \subseteq [-L,L]$ ed esistano M e $\alpha \in [0,1]$ tali che per ogni coppia x, y di punti di (a,b) valga*

$$|f(x) - f(y)| < M|x-y|^\alpha\,. \tag{2.34}$$

Sia $[a',b'] \subseteq (a,b)$. La serie di Fourier di $f(x)$ converge ad $f(x)$ uniformemente *in $[a',b']$.*

Una funzione $f(x)$ ovunque derivabile con derivata limitata,

$$|f'(x)| < M\,,$$

in particolare verifica

$$|f(x) - f(y)| < M|x-y|$$

e quindi soddisfa alle condizioni del teorema. D'altra parte le ipotesi del teorema 94 implicano la continuità della funzione $f(x)$ e questa è una condizione eccessivamente restrittiva per molte applicazioni nelle quali interviene la serie di Fourier. Per cercare di indebolire quest'ipotesi, studiamo cosa accade vicino ad un punto di salto di $f(x)$.

Vale: Vale:

Teorema 95 *La funzione $f(x) \in \mathcal{L}^2(-L,L)$ abbia un salto nel punto x_0. Supponiamo inoltre che esistano $a < x_0$ e $b > x_0$ tali che $f(x)$ ammetta derivata prima continua sia su (a,x_0) che su (x_0,b) e che esistano finiti i limiti*

$$\lim_{x\to x_0-} f'(x)\,, \qquad \lim_{x\to x_0+} f'(x)\,.$$

In questo caso la serie di Fourier di $f(x)$ converge in ogni punto di (a,b) e inoltre:

- *Se $[a',b'] \subseteq (a,x_0)$ oppure se $[a',b'] \subseteq (x_0,b)$ allora la serie* converge uniformemente *ad $f(x)$ su $[a',b']$.*
- *in x_0 la serie di Fourier* converge alla media dei valori dei limiti direzionali in x_0:

$$\frac{1}{2}\left[f(x_0-) + f(x_0+)\right]\,.$$

Esempio 96 Sia

$$\chi(x) = \begin{cases} -1 & \text{se} \quad -\pi < x < 0 \\ 5 & \text{se} \quad x = 0 \\ 1 & \text{se} \quad 0 < x < \pi\,. \end{cases}$$

Si noti che questa funzione differisce dalla funzione $\operatorname{sgn}(x)$ per il valore che assume in 0; ma il valore assunto in un solo punto non altera gli integrali che definiscono i coefficienti di Fourier. Dunque le funzioni definite su $[-\pi, \pi]$ ed uguali a $\chi(x)$ ed a $\operatorname{sgn}(x)$ hanno la medesima serie di Fourier, che è la serie

$$\frac{4}{\pi}\left(\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right)$$

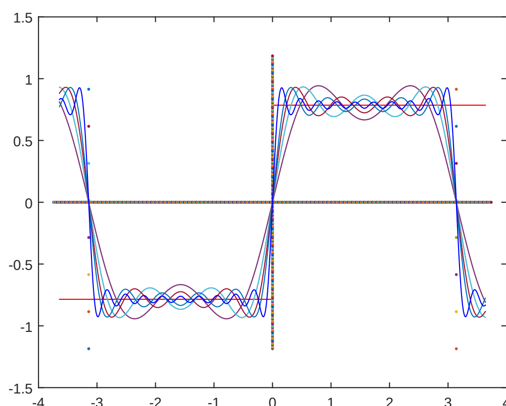
Per $x = 0$ questa serie converge e converge al valore 0, media dei limiti direzionali di $\chi(x)$ per $x \to \pm 0$. Per il teorema 95 la somma della serie è quindi $\operatorname{sgn}(x)$.

La convergenza non può essere uniforme perché le somme parziali sono continue mentre la somma della serie non è continua. Se si disegnano alcune somme parziali, come in figura 2.10, si vede che le somme parziali saltano sopra e sotto il valore ± 1 di una quantità che non si attenua al crescere di N. Calcoli piuttosto laboriosi mostrano che

$$\lim_{N \to +\infty} S_N(1/N) = d$$

e si può mostrare che d è strettamente maggiore di 1: $d > 1{,}089$. E quindi al crescere di N l'errore tra $S_N(x)$ e $\operatorname{sgn}(x)$ non si attenua (ma si concentra sempre di più intorno al salto $x = 0$). ∎

Figura 2.10: Il fenomeno di Gibbs

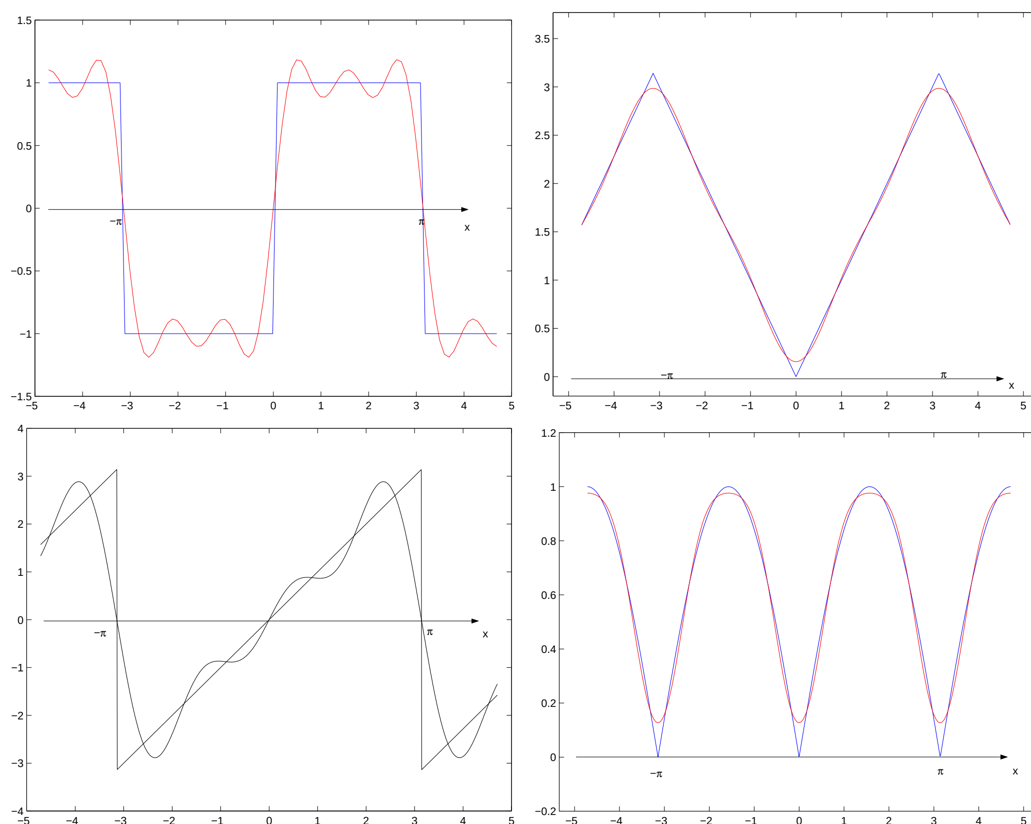


Il fenomeno appena illustrato non dipende dalla particolare funzione $\operatorname{sgn}(x)$ usata nell'esempio. Si può provare che, nelle ipotesi del Teorema 95, esso si verifica in vicinanza di ogni salto. Tale fenomeno va sotto il nome di *Fenomeno di Gibbs.*

Infine, riportiamo alcune serie di Fourier di funzioni $f(x)$ periodiche di periodo 2π e, nelle figure seguenti, i grafici della restrizione della funzione a $(-\pi, \pi)$, con sovrapposti i grafici di alcune somme parziali. Nella colonna di sinistra della tabella, si riporta l'espressione della funzione su $(-\pi, \pi)$. La funzione è poi estesa ad $\mathbf{R}$ per periodicità.

$\operatorname{sign} x$	$\frac{4}{\pi}\left(\frac{\sin x}{1}+\frac{sin 3x}{3}+\frac{\sin 5x}{5}+\cdots\right)$
$\lvert x\rvert$	$\frac{\pi}{2}-\frac{4}{\pi}\left(\frac{\cos x}{1^2}+\frac{\cos 3x}{3^2}+\frac{cos 5x}{5^2}+\cdots\right)$
x	$2\left(\frac{\sin x}{1}-\frac{\sin 2x}{2}+\frac{\sin 3x}{3}-\cdots\right).$
$\begin{cases} x+2\pi & \text{se} \quad -\pi<x<0 \\ x & \text{se} \quad 0<x<\pi \end{cases}$	$\pi-2\left(\frac{\sin x}{1}+\frac{\sin 2x}{2}+\frac{\sin 3x}{3}+\cdots\right)$
$\lvert\sin x\rvert$	$\frac{2}{\pi}-\frac{4}{\pi}\left(\frac{\cos 2x}{1\cdot 3}+\frac{\cos 4x}{3\cdot 5}+\frac{\cos 6x}{5\cdot 7}+\cdots\right).$
$\operatorname{sgn}(x)\cos x$	$\frac{8}{\pi}\left(\frac{\sin 2x}{1\cdot 3}+\frac{2\sin 4x}{3\cdot 5}+\frac{3\sin 6x}{5\cdot 7}\right)$

Figura 2.11: Le serie di Fourier delle restrizioni a $(-\pi, \pi)$ di $f(x) = \operatorname{sign} x$ (sinistra sopra), di $f(x) = |x|$(destra sopra), di $f(x) = x$ (sinistra sotto) e di $f(x) = |\sin x|$(destra sotto)



Capitolo 3

Lo spazio lineare normato $\mathbb{R}^n$

In questo capitolo richiamiamo e precisiamo alcuni concetti che dovrebbero essere noti dai corsi precedenti, in particolare dal corso di Geometria. Tratteremo

- lo spazio lineare $\mathbf{R}^n$.
- introdurremo "norme" e "distanze" in $\mathbf{R}^n$, che permetteranno di definire i punti di accumulazione, gli insiemi aperti e gli insiemi chiusi. Ciò fatto sarà possibile studiare limiti continuità e derivabilità.
- introdurremo quindi il concetto di "insieme convesso" e di "insieme connesso".
- richiameremo le trasformazioni di coordinate in $\mathbf{R}^n$ e la loro relazione con l'orientazione dello spazio.
- introdurremo altri modi (altri "sistemi di coordinate") per rappresentare i punti di $\mathbf{R}^2$ e di $\mathbf{R}^3$.
- studieremo le funzioni da $\mathbf{R}$ in $\mathbf{R}^n$.

I casi su cui insisteremo di più saranno i casi $n = 2$ ed $n = 3$.

3.1 Lo spazio lineare $\mathbb{R}^n$

Non intendiamo qui richiamare la definizione astratta di spazio lineare, vista nei corsi di Geometria. Ci limitiamo a richiamare le sole nozioni necessarie per lo studio di $\mathbf{R}^n$.

Gli elementi di $\mathbf{R}^n$ sono le n-ple ordinate di numeri reali che chiameremo indifferentemente "punti" o "vettori". E' bene essere precisi con la notazione. Una di tali n-ple si potrà scrivere come una sequenza ordinata di n numeri reali $x_1, x_2, \cdots, x_n$ scritta in orizzontale o in verticale, ossia come

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix} \qquad \text{oppure} \qquad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}.$$

Noi useremo sempre la rappresentazione *in verticale.* Però, talvolta sarà conveniente scrivere in orizzontale, con un apice (segno di trasposizione). Ossia intendiamo

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}' = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}.$$

Notiamo inoltre che la definizione di $\mathbf{R}^n$ è suggerita dalla rappresentazione del piano in coordinate cartesiane. Quando $n = 2$ oppure $n = 3$, per indicare il punto $\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}'$ useremo anche la notazione dellla geometria analitica, $P(x_1, x_2, x_3)$.

Gli elementi di $\mathbf{R}^n$, ossia i "vettori", si indicheranno con una lettera in *grassetto:*

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}.$$

Ricordiamo che nei corsi di fisica i vettori si indicano con lettere in grassetto oppure con una freccia sovrapposta, $\vec{x}$; talvolta con una lettera maiuscola, X, oppure con una lettera sottolineata, $\underline{x}$.

I numeri x_i si chiamano le *componenti* del vettore $\mathbf{x}$. Il vettore le cui componenti sono tutte nulle si chiama *vettore nullo* e si indica col simbolo $\mathbf{0}$ (da non confondere col simbolo 0 usato per l'elemento nullo di $\mathbf{R}$). In geometria analitica, il punto $\mathbf{0}$ si chiama anche *origine* delle coordinate e si indica col simbolo O, iniziale di "origine".

Di regola, se $\mathbf{x}$ indica un vettore, le sue componenti si indicheranno come x_i. Ricordiamo che spesso i vettori si indicano col simbolo $\mathbf{v}$, iniziale di "vettore", ma anche col simbolo $\mathbf{r}$, iniziale di "raggio vettore". Questo per quanto riguarda le notazioni con cui si indicano gli elementi di $\mathbf{R}^n$. Lo spazio $\mathbf{R}^n$ però non è solo un isieme di elementi, ma è anche uno *spazio vettoriale*[1]. Infatti gli elementi di $\mathbf{R}^n$ si possono moltiplicare per numeri reali (che si chiamano anche *scalari*) e si possono sommare tra loro. Se α è un numero reale, per definizione si ha

$$\alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \\ \vdots \\ \alpha x_n \end{bmatrix}.$$

[1] si ricordi che uno spazio vettoriale si chiama anche *spazio lineare*.

La somma dei vettori $\mathbf{x}$ e $\mathbf{y}$ si ottiene sommando le componenti corrispondenti: $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}'$, $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_n \end{bmatrix}'$, la loro somma è

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Si rinvia ai corsi di Geometria per le proprietà di queste operazioni. Ricordiamo però che se $n = 3$ l'operazione di somma appena definita corrisponde alla somma di vettori con la regola del parallelogramma nota dai corsi di Fisica.

Come terminologia, diremo anche che $\mathbf{x} + \mathbf{y}$ è ottenuto traslando di $\mathbf{x}$ il vettore $\mathbf{y}$ o, in modo equivalente che è ottenuto traslando di $\mathbf{y}$ il vettore $\mathbf{x}$.

Si considerino ora r vettori $\mathbf{v}_1, \ldots, \mathbf{v}_r$ di $\mathbf{R}^n$. Il vettore

$$\sum_{i=1}^{r} \alpha_i \mathbf{v}_i$$

(con α_i numeri reali) si chiama *combinazione lineare* dei vettori $\mathbf{v}_i$.

L'insieme delle combinazioni lineari si chiama *s.spazio lineare generato dai vettori* $\mathbf{v}_1$, ..., $\mathbf{v}_r$ *di* $\mathbf{R}^n$. Esso contiene sempre $\mathbf{0}$, che si ottiene scegliendo nulli tutti gli α_i. Si potrebbe trovare $\mathbf{0}$ anche con altre scelte degli α_i. Se ciò non accade, i vettori $\mathbf{v}_i$ si chiamano *linearmente indipendenti*. Ossia, i vettori $\mathbf{v}_i$ sono linearmente indipendenti quando

$$\sum_{i=1}^{r} \alpha_i \mathbf{v}_i = \mathbf{0}$$

implica $\alpha_i = 0$ per tutti gli indici i. In questo caso l'insieme dei vettori $\mathbf{v}_i$ si chiama una *base* del s.spazio. Se $r = n$ e se i vettori $\mathbf{v}_i$ sono linearmente indipendenti, lo spazio da essi generato è $\mathbf{R}^n$ e si dice che i vettori $\mathbf{v}_i$ sono una base di $\mathbf{R}^n$.

Osserviamo che ogni s.spazio lineare è esso stesso uno spazio lineare rispetto alle operazioni di somma e di prodotto per scalari.

Come *base* di $\mathbf{R}^n$ si potranno scegliere n vettori indipendenti qualsiasi. Scegliendo però i vettori $\mathbf{e}_k$,

$$\mathbf{e}_k = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

con 1 nella posizione k e gli altri elementi tutti nulli, si ha la *base canonica*.

Esempio 97 Si sa dalla geometria analitica che i s.spazi di $\mathbf{R}^2$ sono $\mathbf{0}$, $\mathbf{R}^2$ e le rette per l'origine. Se invece $n = 3$ i s.spazi sono $\mathbf{0}$, $\mathbf{R}^3$, le rette per l'origine ed i piani per l'origine. ∎

Siano ora X ed Y due s.insiemi di $\mathbf{R}^n$. Col simbolo $X+Y$ si intende l'insieme

$$X+Y=\{\mathbf{x}+\mathbf{y}\,,\ \mathbf{x}\in X\,,\ \mathbf{y}\in\ Y\}\,.$$

A noi interessa in particolare il caso in cui Y è un s.spazio mentre X ha l'unico elemento $\mathbf{x}_0$. L'insieme $X+Y$ in questo caso si indica col simbolo

$$\mathbf{x}_0+Y$$

e si chiama *s.spazio* *affine* di $\mathbf{R}^n$, parallelo ad Y, ottenuto traslando Y *di* $\mathbf{x}_0$ o anche *in* $\mathbf{x}_0$. Si noti che:

- Se $\mathbf{x}_0\notin Y$ allora $\mathbf{0}\notin\mathbf{x}_0+Y$.
- in generale, un s.spazio affine non è un s.spazio lineare. Lo è se e solo se si ha $\mathbf{x}_0+Y=Y$ e ciò avviene se e solo se $\mathbf{x}_0\in Y$;

Come terminologia, in generale chiameremo semplicemente "s.spazi" i s.spazi lineari mentre l'aggettivo "affine" non verrà mai omesso, salvo nei casi particolari seguenti, nei quali useremo i termini "retta" e "piano" propri della geometria:

- Gli spazi della forma $\{t\mathbf{y}_0\}_{t\in\mathbf{R}}$ (con $\mathbf{y}_0\neq 0$) si chiamano *rette per l'origine* e $\mathbf{x}_0+t\mathbf{y}_0$, $t\in\mathbf{R}$, si chiama *retta per* $\mathbf{x}_0$ parallela ad $\mathbf{y}_0$ (che deve essere non nullo).
- si chiama *piano per l'origine* *in* $\mathbf{R}^n$ l'insieme dei punti $\mathbf{x}$ le cui componenti x_1, ...,x_n verificano l'equazione lineare

  $$a_1x_1+a_2x_2+\cdot+a_nx_n=0\,.$$

  I parametri a_1, ... a_n sono fissati, non tutti nulli.
  Se X è un piano per l'origine ed $\mathbf{y}=\begin{bmatrix} y_1 & \ldots & y_n\end{bmatrix}'$ un punto fissato di $\mathbf{R}^n$, l'insieme $\mathbf{y}+X$ si chiama *piano per* $\mathbf{y}$ e si vede facilmente che le componenti dei suoi punti verificano

  $$a_1(x_1-y_1)+a_2(x_2-y_2)+\cdot+a_n(x_n-y_n)=0\,.$$

Vettori colineari e rette parallele

Due vettori non nulli $\mathbf{v}$ e $\mathbf{w}$ si dicono *colineari* quando le componenti corrispondenti sono proporzionali, ossia quando esiste $\alpha\in\mathbf{R}$, $\alpha\neq 0$, tale che

$$\mathbf{v}=\alpha\mathbf{w}\,.$$

Col linguaggio della geometria analitica, i due vettori $\mathbf{v}$ e $\mathbf{w}$ sono colineari quando identificano la medesima retta uscente dall'origine.

Consideriamo ora due rette,

$$\{\mathbf{x}_0+t\mathbf{v}\,,\ t\in\mathbf{R}\}\,,\qquad \{\mathbf{y}_0+\tau\mathbf{w}\,,\ \tau\in\mathbf{R}\}\,. \tag{3.1}$$

Le due rette si dicono *parallele* quando i due vettori $\mathbf{v}$ e $\mathbf{w}$ sono colineari. Geometricamente questo significa che un punto Q della seconda retta si ottiene da un punto P della prima retta, traslandolo mediante il vettore $\mathbf{y}_0-\mathbf{x}_0$, indipendente dai punti P e Q considerati.

Due rette non parallele possono avere un punto comune, o meno. Nel secondo caso le rette si dicono *sghembe*. Se le due rette hanno un punto comune, si dicono *incidenti*. Ciò accade quando esistono t e τ tali che

$$\mathbf{x}_0 + t\mathbf{v} = \mathbf{y}_0 + \tau\mathbf{w} \qquad \text{ossia} \qquad \mathbf{x}_0 - \mathbf{y}_0 = \tau\mathbf{w} - t\mathbf{v}\,.$$

Dunque, le due rette (3.1) sono incidenti se e solo se $\mathbf{x}_0 - \mathbf{y}_0$ appartiene al piano generato da $\mathbf{v}$ e $\mathbf{w}$.

Osservazione 98 *Sia $m \neq 0$. Le due rette*

$$\mathbf{x}_0 + t\mathbf{y}_0\,, \qquad \mathbf{x}_0 + tm\mathbf{y}_0\,, \qquad t \in \mathbf{R}$$

coincidono. ∎

3.1.1 Connessione e convessità

I punti di una retta hanno rappresentazione

$$\mathbf{x}_0 + t\mathbf{y}_0\,, \qquad t \in \mathbf{R}\,. \tag{3.2}$$

Il punto $\mathbf{x}_0$ si ritrova scegliendo $t = 0$. Per questo, come si è detto, questa retta si chiama "retta per $\mathbf{x}_0$, parallela ad $\mathbf{y}_0$".

Sia ora $\mathbf{x}_1$ un secondo punto di $\mathbf{R}^n$. Chiediamoci se, per qualche selta di $\mathbf{y}$, la retta (3.2) contenga $\mathbf{x}_1$ ossia, come si dice, "passi anche per $\mathbf{x}_1$".

Ciò avviene quando per un certo valore t_1 del "parametro" t si ha

$$\mathbf{x}_0 + t_1\mathbf{y} = \mathbf{x}_1\,.$$

Questo è un insieme di n equazioni nelle n componenti di $\mathbf{y}$. L'uguaglianza si ottiene solo quando $\mathbf{y}$ è dato da

$$\mathbf{y} = \frac{1}{t_1}\left[\mathbf{x}_1 - \mathbf{x}_0\right]\,.$$

Il valore t_1 può scegliersi arbitrariamente, purché non nullo[2]. Dunque, scelto $t_1 = 1$, la retta per $\mathbf{x}_0$ e $\mathbf{x}_1$ si rappresenta come

$$\mathbf{x} = \mathbf{x}_0 + t\left[\mathbf{x}_1 - \mathbf{x}_0\right]\,.$$

Il suo s.insieme

$$\{\mathbf{x}_0 + t\left[\mathbf{x}_1 - \mathbf{x}_0\right]\,, \quad t \in [0,1]\}$$

si chiama *segmento* congiungente $\mathbf{x}_0$ ed $\mathbf{x}_1$. I punti $\mathbf{x}_0$ ed $\mathbf{x}_1$ si chiamano gli *estremi* del segmento. Precisamente, $\mathbf{x}_0$, ottenuto per $t = 0$, si chiama il "primo estremo" ed $\mathbf{x}_1$, ottenuto per $t = 1$, si chiama il "secondo estremo"[3].

Se $\mathbf{x}_0 = \mathbf{x}_1$, il segmento degenera nel punto $\mathbf{x}_0$:

$$\mathbf{x}_0 = \mathbf{x}_0 + t\left[\mathbf{x}_1 - \mathbf{x}_0\right] = \mathbf{x}_0 + t\left[\mathbf{x}_0 - \mathbf{x}_0\right]$$

---

[2]perchè, come si è già notato, la retta per $\mathbf{x}_0$ parallela ad $\mathbf{y}_0$ non muta sostituendo $\mathbf{y}_0$ con $m\mathbf{y}_0$, purché sia $m \neq 0$.

[3]si osservi che la sostituzione $t = 1 - \tau$, $\tau \in (0,1)$, scambia il primo col secondo estremo.

per ogni t.

Definiamo ora cosa si intende per "spezzata" di $\mathbf{R}^n$. Questo termine indica un numero finito di segmenti che "si susseguono"; ossia tali che il secondo estremo di uno sia anche primo estremo del successivo; ossia, consideriamo un numero finito di segmenti I_1, I_2, ..., I_k. Se accade che per $1 < j < k$ il primo estremo di I_{j+1} coincide col secondo estremo di I_j, l'insieme $\cup_{j=1}^k I_j$ si dice una *spezzata* .

Si dice che una spezzata $\cup_{j=1}^k I_j$ congiunge i due punti $\mathbf{x}$ ed $\mathbf{y}$ quando $\mathbf{x}$ è il primo estremo di I_1 e $\mathbf{y}$ è il secondo estremo di I_k. Se accade che tali due punti coincidono, la spezzata si dice *chiusa* .

Sia ora A un s.insieme di $\mathbf{R}^n$. L'insieme A si dice *connesso* [4] quando ogni coppia di punti di A può essere congiunta con una spezzata i cui punti appartengono ad A.

Inoltre, si considerano connessi anche gli insiemi costituiti da un solo punto.

Un insieme che è sia aperto che connesso si chiama un *dominio*.

Può accadere che l'insieme connesso A contenga un punto $\mathbf{x}_0$ con questa proprietà: ogni altro punto $\mathbf{x}_1 \in A$ può essere congiunto ad $\mathbf{x}_0$ mediante un segmento contenuto in A; ossia mediante una spezzata costituita da un unico segmento. In tal caso l'insieme A si dice *stellato* *rispetto ad* $\mathbf{x}_0$.

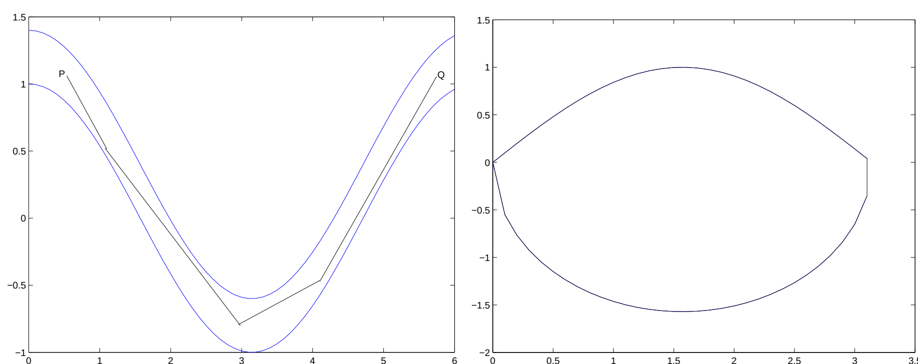
Sia ora C un s.insieme di $\mathbf{R}^n$. Si dice che C è *convesso*

- quando è vuoto oppure contiene un solo punto
- oppure quando contiene il segmento congiungente due qualsiasi dei suoi punti.

Ossia, se a C appartengono almeno due punti, allora C è convesso quando è stellato rispetto a ciascuno dei suoi punti.

La figura 3.1 rappresenta un insieme connesso, a sinistra, ed un insieme convesso, a destra.

Figura 3.1: Insieme connesso, a sinistra, e convesso a destra.



[4]più precisamente si dovrebbe dire "connesso per archi". La definizione che qui diamo non ha la forma più generale possibile. Però noi saremo principalmente interessati ad insiemi connessi che sono anche "aperti" e per tali insiemi la definizione data coincide con quella generale, che non riportiamo.

Per $n = 1$ le tre definizioni di insieme connesso per archi, stellato e convesso si riducono a quella di intervallo.

3.1.2 Vettori liberi e vettori applicati

L'uso dei vettori è suggerito dalle applicazioni fisiche: un "vettore" può rappresentare, per esempio, una forza o uno spostamento. Domanda ovvia: spostamento da dove, o forza applicata dove? I vettori come n-ple ordinate di numeri reali non permettono di rispondere a queste domande. Tali vettori possono usarsi per rappresentare uno "spostamento" nel senso della distanza percorsa, in una certa direzione e in un certo verso, indipendentemente da quale sia il punto di partenza; o una forza di una certa intensità diretta secondo una certa direzione e con un certo verso, indipendentemente da dove essa sia applicata. Per questa ragione, i vettori che abbiamo introdotto si chiamano in fisica *vettori liberi*. Si può decidere di applicare tutti i vettori liberi in un punto convenzionalmente scelto. La scelta naturale è di applicarli nell'origine: i vettori liberi verranno interpretati anche come vettori applicati nell'origine e quindi, per esempio, il vettore $\begin{bmatrix} 1 & 3 & 5 \end{bmatrix}'$ rappresenta lo spostamento, in linea retta, da O al punto $P(1, 3, 5)$.

Se vogliamo rappresentare un *vettore applicato* dobbiamo dare una coppia di vettori: il primo rappresenta il punto di applicazione e il secondo rappresenta il vettore (forza, spostamento,...) in esso applicato. A noi non serve essere troppo formali a questo proposito, ma è necessario sapere che:

- sui vettori applicati nel medesimo punto si fanno tutte le operazioni (sia quelle già introdotte che quelle che introdurremo) che possono farsi tra vettori liberi. Dunque, se $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}'$ e $\begin{bmatrix} 4 & 5 & 6 \end{bmatrix}'$ rappresentano due vettori applicati nel medesimo punto, per esempio nel punto $P(4, 4, 4)$, la loro somma è il vettore $\begin{bmatrix} 5 & 7 & 9 \end{bmatrix}'$ ancora applicato in $P(4, 4, 4)$.
- **Non si fanno operazioni tra vettori applicati in punti diversi.**
- Un vettore $\mathbf{v}$ applicato in P può *spostarsi per parallelismo* in un vettore applicato in Q procedendo come segue: al vettore $\mathbf{v}$ applicato in P si fa corrispondere il vettore $\mathbf{v}$ applicato in Q.
- Se si vogliono fare operazioni tra vettori applicati in punti diversi, bisogna prima di tutto traslarli per parallelismo, applicandoli in un punto comune.

3.2 Basi e basi ordinate

Ricordiamo che una base di $\mathbf{R}^n$ è un insieme di n vettori linearmente indipendenti. Un insieme non cambia se si cambia l'ordine col quale se ne elencano gli elementi. Se però B è un insieme finito e si stabilisce un ordine tra i suoi elementi, si dice che B è un insieme ordinato e se B è una base di $\mathbf{R}^n$, si dice che B è una *base ordinata*. Per esempio, se si stabilisce di elencare gli elementi della base canonica elencando e_j al j–mo posto si ha una base ordinata, ma si ha una base ordinata anche se si stabilisce di elencarne gli elementi a rovescio, oppure prima quelli di indice pari e poi quelli di indice dispari.

Quando in $\mathbf{R}^n$ si è stabilita una base ordinata si possono fare cose che non sono possibili con basi non ordinate. Per esempio, sia $\mathcal{A}$ è una trasformazione lineare da $\mathbf{R}^n$ in $\mathbf{R}^m$ e

siano rispettivamente $\{\mathbf{e}_i\}$ ed $\{\tilde{\mathbf{e}}_j\}$ due basi ordinate la prima di $\mathbf{R}^n$ e la seconda di $\mathbf{R}^m$. La trasformazione $\mathcal{A}$ si può rappresentare con una matrice come segue. Si considera l'elemento $\mathcal{A}\mathbf{e}_1 \in \mathbf{R}^m$. Questo si rappresenta come

$$\mathcal{A}\mathbf{e}_1 = \sum_{j=1}^{m} a_1^j \tilde{\mathbf{e}}_j \,.$$

Si costruisce una matrice mettendo il numero a_1^j nella posizione j della prima colonna. La seconda colonna si costruisce in modo analogo a partire da $\mathcal{A}\mathbf{e}_2$ e così via fino a costruire l'n-ma colonna.

Si costruisce così una matrice A che rappresenta la trasformazione lineare $\mathcal{A}$.

Se la base prescelta è quella canonica e se non si stabilisce diversamente in modo esplicito, si intende che la base canonica è una base ordinata e che gli elementi della base si susseguono nell'ordine dei loro indici:

$$\mathbf{e}_1\,,\mathbf{e}_2\,,\ldots\,,\mathbf{e}_n\,.$$

Si dice che *una base ordinata subordina un' orientazione di* $\mathbf{R}^n$. Visto che una base di $\mathbf{R}^n$ contiene n elementi, ci sono $n!$ modi di elencarli e quindi si potrebbe pensare che in $\mathbf{R}^n$ ci siano (almeno) $n!$ orientazioni diverse. Invece non è così. Consideriamo per questo una base ordinata B_1 ed una seconda base ordinata B_2, (che potrebbe essere ottenuta dagli stessi elementi di B_1, ordinati in modo diverso). Si sa dal corso di Geometria che i cambiamenti di base si rappresentano mediante una matrice invertibile. Sia P la matrice che trasforma ordinatamente gli elementi di B_1 in quelli di B_2. Essendo P invertibile, il suo determinante non è zero e quindi delle due l'una:

$$\det P > 0 \quad \text{oppure} \quad \det P < 0\,.$$

Se $\det P > 0$ si dice che le due basi B_1 e B_2 subordinano la medesima orientazione di $\mathbf{R}^n$, altrimenti subordinano orientazioni opposte.

Dunque, in $\mathbf{R}^n$ si trovano due orientazioni, che si dicono opposte l'una dell'altra.

3.2.1 Il piano e lo spazio

I punti del piano si mettono in corrispondenza biunivoca con quelli di $\mathbf{R}^2$ procedendo come segue: si fissano due rette incidenti (e tra loro diverse) r ed s del piano ed un'unità di misura per le lunghezze[5]. Il punto comune alle due rette si chiama *origine*. Su ciascuna delle due rette si fissa un verso (che si chiama "positivo"). Il segmento PO della retta r ha lunghezza positiva quando un punto che parte da O incontra P muovendosi nel verso positivo; negativo altrimenti. In tal caso si dice che P appartiene al semiasse positivo. La stessa convenzione si usa sulla retta s.

Si fissa quindi un ordine tra le due rette. La prima si chiama asse delle *ascisse* o *asse x* e la seconda asse delle *ordinate* o *asse y*. Le due rette si chiamano *assi coordinati*. Le rette del piano parallele agli assi coordinati si chiamano *rette coordinate*.

[5]che assumeremo la medesima sullle due rette, ma si potrebbero anche fissare unità di misura diverse, una sulla prima e una sulla seconda retta.

Facendo ciò, si dice che si è definita un' *orientazione* del piano. Si fa quindi corrispondere l'orientazione del piano con quella di $\mathbf{R}^2$ associando il vettore $\mathbf{e}_1$ al punto dell'asse delle ascisse a distanza $+1$ dall'origine e il vettore $\mathbf{e}_2$ col punto dell'asse delle ordinate a distanza $+1$ dall'origine.

Sia P un punto del piano. Si fanno passare per P due rette r_1, parallela all'asse delle ordinate, ed r_2 parallela a quella delle ascisse. Sia P_1 il punto in cui r_1 incontra l'asse delle ascisse e P_2 l'intersezione di di r_2 con quello delle ordinate. Siano x ed y le lunghezze, con segno, rispettivamente di OP_1 e di OP_2. Al punto P si fa corrispondere il vettore

$$\mathbf{x} = x\mathbf{e}_1 + y\mathbf{e}_2 .$$

Viceversa, ad ogni vettore $\mathbf{x}$ si fa corrispondere un punto del piano.

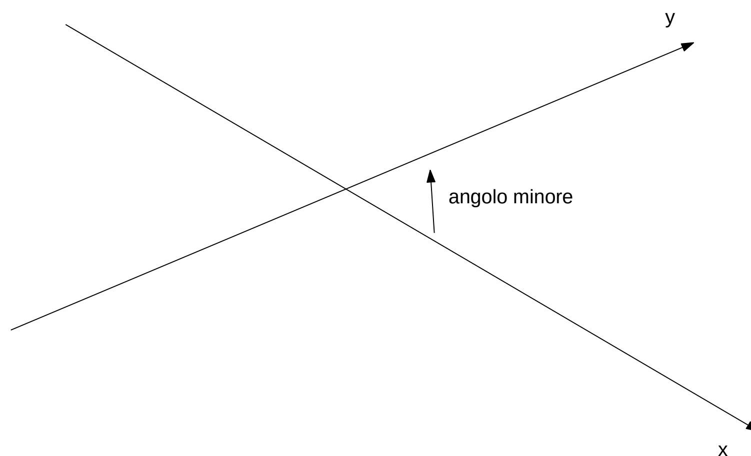
Si noti che perchè ciò abbia senso, va stabilito prima quale asse scegliere come asse delle ascisse, e quale dei due elementi della base considerare per primo. Ossia, la corrispondenza biunivoca che abbiamo costruito è tra il piano, a cui abbiamo imposto un'orientazione, ed $\mathbf{R}^2$, a cui abbiamo imposto un'orientazione.

In pratica, si segue questa convenzione:

- La base canonica di $\mathbf{R}^2$ si ordina scegliendo prima $\mathbf{e}_1$;
- sia fissato l'asse delle ascisse e il verso su di esso. Il semiasse "positivo" delle ascisse si può sovrapporre ad uno dei semiassi delle ordinate con due rotazioni una oraria ed una antioraria. Dei due angoli, uno è minore dell'altro. L'orientazione positiva sull'asse delle ordinate si sceglie in modo che la sovrapposizione avvenga girando l'asse delle ascisse in verso antiorario e dell'angolo minore.

La figura 3.2 illustra la situazione.

Figura 3.2: Senso positivo di rotazione e verso sugli assi coordinati



Il più delle volte gli assi coordinati si prendono ortogonali tra loro e l'asse delle ascisse si rappresenta orizzontale, col verso positivo verso destra. In tal caso il verso positivo dell'asse delle ordinate punta verso l'alto.

La rappresentazione dello spazio è analoga. Senza entrare nei dettagli, consideriamo subito il caso di un sistema cartesiano ortogonale. Si scelgono tre rette tra loro ortogonali che si chiamano rispettivamente *delle ascisse* o *asse x*, delle *ordinate* o *asse y*, delle *quote* o *asse z*. Gli assi x ed y identificano un piano, il piano (x, y), a cui l'asse z è ortogonale. L'orientazione sul piano (x, y) si fissa come si è detto sopra. Rimane quindi da scegliere il verso positivo dell'asse z. Questo si sceglie in modo che un osservatore in piedi sul piano (x, y), appoggiato all'asse z e con la testa nel verso positivo veda che il semiasse x positivo si riporta sul semiasse y positivo ruotandolo dell'angolo minore e in verso antiorario.

Consideriamo ora il punto $(1, 0, 0)$. Ruotando l'asse delle ascisse di un'angolo giro, esso descrive una circonferenza e, se l'orientazione del piano (x, y) è positiva, raggiunge il punto $(0, 1, 0)$ dopo una rotazione di un angolo retto in verso antiorario. Un insetto che parta da $(1, 0, 0)$ e si muova lungo la circonferenza verso il punto $(0, 1, 0)$, raggiungendolo dopo la rotazione di $\pi/2$, vede il disco alla sua sinistra.

E' questa la prima comparsa della regola *d'Ampère* per la determinazione dell'orientazione dello spazio.

Ogni punto P si rappresenta mediante le sue tre coordinate x, y e z. Queste si ottengono facendo passare per P tre piani, paralleli ai piani individuati dalle coppie di assi coordinati (che si chiamano *piani coordinati*). L'ascissa x di P è la distanza dall'origine dell'intersezione tra l'asse delle ascisse e il piano per P paralello agli assi y e z, presa con segno. L'ordinata e la quota si definiscono in modo analogo.

Si chiamano *rette coordinate* le rette dello spazio parallele agli assi cartesiani.

Sia ora P un punto (del piano o dello spazio), $P(x, y, x)$. Facciamogli corrispondere il vettore $\mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ che si interpreta come "spostamento" percorso da un punto che partendo dall'origine raggiunge la posizione occupata da P. In questo modo, lo spazio si mette in corrispondenza biunivoca con $\mathbf{R}^3$, che si pensa orientato mediante la sua base canonica.

Quando si lavora con un sistema di assi cartesiani ortogonali, si usano i simboli sequenti: lavorando sul piano,

$$\mathbf{i} = \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{j} = \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Invece lavorando nello spazio

$$\mathbf{i} = \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{j} = \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{k} = \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

3.3 Norme e distanze

La teoria dei limiti per le funzioni di una variabile dipende in modo essenziale dalle proprietà seguenti del valore assoluto:

- Per ogni x reale, $|x| \geq 0$ e $|x| = 0$ se e solo se $x = 0$;
- Il valore assoluto di un prodotto è il prodotto dei valori assoluti: $|xy| = |x| \cdot |y|$;
- la disuguaglianza triangolare: $|x + y| \leq |x| + |y|$.

Ricordiamo che da queste proprietà segue anche:

$$\Big| |x| - |y| \Big| \le |x - y| \,.$$

Esaminando il corso di Analisi Matematica 1, si vede facilmente che queste sono le uniche proprietà che servono per la teoria dei limiti, eccezion fatta per quei teoremi che richiedono una relazione di ordine, come i teoremi del confronto o delle funzioni monotone.

L'osservazione precedente suggerisce di definire *norma* su $\mathbf{R}^n$ una funzione definita su $\mathbf{R}^n$ ed a valori reali, con le proprietà che ora descriviamo. Una norma si indica col simbolo $\|\cdot\|$. Con questo simbolo, le proprietà sono:

- la norma prende valori non negativi: $\|\mathbf{x}\| \ge 0$ per ogni $\mathbf{x}$;
- la norma si annulla solo in $\mathbf{0}$: $\|\mathbf{x}\| = 0$ implica $\mathbf{x} = \mathbf{0}$;
- vale $\|t\mathbf{x}\| = |t| \cdot \|\mathbf{x}\|$ per ogni $\mathbf{x}$ e per ogni numero reale t. Si noti che scegliendo $t = 0$ si trova $\|\mathbf{0}\| = 0$.
- vale la disuguaglianza triangolare: per ogni coppia di vettori $\mathbf{x}$ e $\mathbf{y}$ si ha:

$$\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\| \,.$$

Non è difficile provare:

Lemma 99 *Per ogni* $\mathbf{x}$, $\mathbf{y}$ *vale*

$$\Big| \|\mathbf{x}\| - \|\mathbf{y}\| \Big| \le \|\mathbf{x} - \mathbf{y}\| \,.$$

Dim. Esattamente come nel caso dei numeri, si nota che la disuguaglianza da provare equivale alle due disuguaglianze

$$-\|\mathbf{x} - \mathbf{y}\| \le \|\mathbf{x}\| - \|\mathbf{y}\| \le \|\mathbf{x} - \mathbf{y}\|.$$

La seconda disuguaglianza segue dalla disuguaglianza triangolare, scrivendo

$$\|\mathbf{x}\| = \|\mathbf{x} - \mathbf{y} + \mathbf{y}\| \le \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\| \,.$$

L'altra segue in modo analogo, scrivendo

$$\|\mathbf{y}\| = \|\mathbf{y} - \mathbf{x} + \mathbf{x}\| \le \|\mathbf{y} - \mathbf{x}\| + \|\mathbf{x}\| = \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{x}\| \,. \quad \blacksquare$$

Dunque, definita una norma, è possibile trattare la teoria dei limiti su $\mathbf{R}^n$ esattamente come per $n = 1$, provando tutti i medesimi teoremi, con le stesse dimostrazioni, a parte quelli che fanno intervenire la monotonia.

Mostriamo che norme su $\mathbf{R}^n$ esistono:

Esempio 100 Le seguenti sono norme su $\mathbf{R}^2$:

$$\left\| \begin{bmatrix} x & y \end{bmatrix}' \right\|_1 = |x| + |y| \,, \qquad \left\| \begin{bmatrix} x & y \end{bmatrix}' \right\|_\infty = \max\{|x|, |y|\} \,. \quad \blacksquare$$

Si mostri per esercizio che le due funzioni definite sopra effettivamente soddisfano alle proprietà richieste per la definizione di norma. ∎

Quindi, su $\mathbf{R}^2$ possono definirsi almeno due norme diverse. In realtà si possono definire infinite norme diverse. Infatti:

Teorema 101 *Sia $p \geq 1$ e sia $\mathbf{R}^n$ rappresentato rispetto alla base canonica. Ciascuna delle seguenti è una norma su $\mathbf{R}^n$:*

$$\|\mathbf{x}\|_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{1/p} .$$

Diciamo subito che la norma di gran lunga più importante è la norma che corrisponde al numero $p = 2$, che si chiama *norma euclidea*:[6]

$$\|\mathbf{x}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} .$$

Introdotta una norma si può introdurre la *distanza* tra i vettori,

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

e quindi definire:

Definitione 1 *Sia $\mathbf{x_0}$ un punto di $\mathbf{R}^n$ e sia $r > 0$. Si chiama intorno di $\mathbf{x_0}$ di raggio r, o palla aperta di centro $\mathbf{x_0}$ e raggio r l'insieme*

$$B(\mathbf{x_0}, r) = \{\mathbf{v} \mid \|\mathbf{x_0} - \mathbf{v}\| < r\} .$$

Dato un insieme A di $\mathbf{R}^n$ ed un vettore $\mathbf{x_0}$, si dice che:

- *Un insieme è limitato quando esiste una palla che lo contiene.*
- *$\mathbf{x_0}$ è interno ad A se esiste $r > 0$ tale che $B(\mathbf{x_0}, r) \subseteq A$;*
- *Si dice che $\mathbf{x_0}$ è punto di accumulazione per A se* per ogni $r > 0$ esiste $\mathbf{a} \in A$, *$\mathbf{a} \neq \mathbf{x_0}$, con $\mathbf{a} \in B(\mathbf{x_0}, r)$.*
- *Il punto x_0 è punto della frontiera di A se non è interno né ad A né al suo complementare.*
- *Un insieme si dice aperto se tutti i suoi punti sono interni, oppure se è vuoto; chiuso se contiene tutti i suoi punti di accumulazione oppure se è vuoto.*

  Si mostra facilmente che un insieme è aperto se e solo se il suo complementare è chiuso.
- *Si chiama successione a valori in $\mathbf{R}^n$ una funzione che ad ogni numero naturale associa un vettore di $\mathbf{R}^n$.*
- *Una successione è limitata quando limitata è la sua immagine.*

---

[6]si osservi che i valori assoluti nell'espressione seguente non hanno alcun ruolo. Sono stati introdotti per due ragioni: prima di tutto per uniformità col caso $p \neq 2$ e poi perché, come diremo in seguito, niente cambia se i vettori si considerano a componenti complesse, invece che reali; salvo che in tal caso anche la definizione di $\|\mathbf{x}\|_2$ richiede i moduli perché la norma deve comunque essere un numero reale.

- *Sia* $\{\mathbf{v}_n\}$ *una successione. Si dice che*

$$\lim_n \mathbf{v}_n = \mathbf{v}_0$$

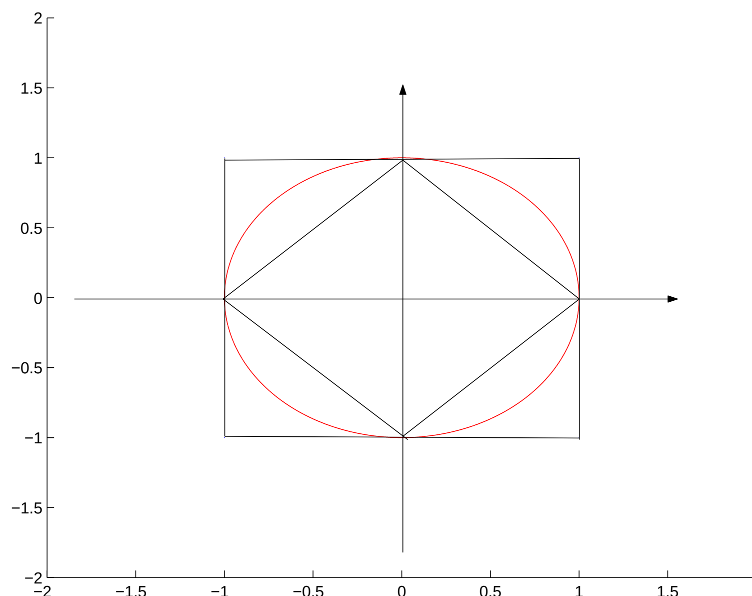
quando, per ogni $\epsilon > 0$ *esiste* N_ϵ *tale che se* $n > N_\epsilon$ *si ha* $\mathbf{v}_n \in B(\mathbf{v}_0, \epsilon)$.

Dato che le norme su $\mathbf{R}^n$ sono infinite, potrebbe sembrare che ci siano infinite teorie dei limiti tra loro diverse. Invece, fortunatamente, le proprietà di avere o non avere limite, il valore dell'evenuale limite e le proprietà dei limiti non dipendono dalla definizione di norma che si decide di usare in $\mathbf{R}^n$. Prima di provare questo fatto, illustriamone la ragione intuitiva consideriamo le tre norme $\|\cdot\|_1$, $\|\cdot\|_2$ e $\|\cdot\|_\infty$ in $\mathbf{R}^2$.

La figura 3.3 mostra le palle di centro 0 e raggio 1 relative alle tre norme. La palla relativa alla norma euclidea è un disco, in figura rappresentata come un ellisse perché l'unità di misura sui due assi non è la medesima. Le altre due palle sono quadrati (rettangoli in figura, per la ragione detta sopra). Quello con i lati paralleli agli assi coordinati è la palla nella norma $\|\cdot\|_\infty$.

Si vede da questa figura che se una successione $\mathbf{v}_n$ tende a zero rispetto ad una di queste norme, entra e rimane definitivamente dentro ciascuna delle tre palle; e quindi tende a zero anche rispetto alle altre norme.

Figura 3.3: Sfera di centro l'origine in norme diverse



Il risultato generale è conseguenza delle disuguaglianze seguenti. La prima è ovvia mentre omettiamo la dimostrazione della seconda.

Lemma 102 *Sia* $\|\cdot\|_p$, $1 \le p \le +\infty$ *una norma di* $\mathbf{R}^n$. *Valgono le due disuguaglianze seguenti:*

- *Per ogni* i *vale* $|x_i| \le \|\mathbf{x}\|$;
- *Esiste un numero* M *(che dipende da* n *e da* p*) tale che* $\|\mathbf{x}\|_p \le M\sum_{i-1}^n |x_i| = \|\mathbf{x}\|_1$.

In definitiva, per $1 \le p \le +\infty$ *si ha*

$$\|\mathbf{x}\|_1 \le n\|\mathbf{x}\|_p \le Mn\|\mathbf{x}\|_1 \,. \tag{3.3}$$

Accettando questo lemma possiamo provare:

Teorema 103 *In* $\mathbf{R}^n$, *una successione* $(\mathbf{v}_n)$ *converge a* $\mathbf{v}_0$ *in norma* p *se e solo se converge al medesimo limite* $\mathbf{v}_0$ *in qualsiasi altra norma* q. *Si ha* $\|\mathbf{v}_n\|_p \to +\infty$ *se e solo se* $\|\mathbf{v}_n\|_q \to +\infty$.

Dim. Proviamo il primo asserto. Sostituendo $\mathbf{v}_n$ con $\mathbf{v}_n - \mathbf{v}_0$ si può studiare il caso della convergenza a zero. Proviamo che $\|\mathbf{v}_n\|_p \to 0$ se e solo se $\|\mathbf{v}_n\|_1 \to 0$. Questo è immediato dal teorema del confronto per i limiti (di successioni a valori reali) applicato alla disuguaglianza (3.3).

Il secondo asserto si prova in modo ovvio. ∎

Inoltre, sia (v_n^i) la successione di numeri reali ottenuta considerando la componente i-ma dei vettori $\mathbf{v}_n$. Le disuguaglianze (3.3) mostrano:

Teorema 104 *La successione* $(\mathbf{v}_n)$ *converge se e solo se ciascuna delle successioni di numeri reali* (v_n^i) *è convergente e il vettore* $\lim \mathbf{v}_n$ *ha per i-ma componente il numero* $\lim v_n^i$.

Osservazione 105 Si noti che un asserto analogo non vale per *successioni divergenti* ossia tali che

$$\lim \|\mathbf{v}_n\| = +\infty \,.$$

Per questo basta che una delle successioni (v_n^i) diverga! ∎

3.3.1 Completezza di $\mathbf{R}^n$

Si chiama *successione fondamentale* o *successione di Cauchy* una successione $(\mathbf{v}_n)$ con questa proprietà:

Per ogni $\epsilon > 0$ esiste N_ϵ tale che se $n > N_\epsilon$, $m > N_\epsilon$ allora

$$\|\mathbf{v}_n - \mathbf{v}_m\| < \epsilon \,.$$

In simboli,

$$\forall \epsilon > 0 \; \exists N_\epsilon \mid n > N_\epsilon \,, \; m > N_\epsilon \Longrightarrow \|\mathbf{v}_n - \mathbf{v}_m\| < \epsilon \,.$$

Una dimostrazione analoga a quella del Teorema 103 mostra che la proprietà di essere fondamentale non dipende dalla norma usata e per questo nella definizione precedente abbiamo usato il generico simbolo di norma. Inoltre le disuguaglianze (3.3) mostrano che:

Teorema 106 *La successione* $(\mathbf{v}_n)$ *è fondamentale se e solo se ciascuna delle sue componenti è una successione fondamentale di numeri reali.*

Si sa dal corso di Analisi Matematica 1 che una successione di numeri converge se e solo se è fondamentale. E quindi quest'asserto vale anche in $\mathbf{R}^n$:

Teorema 107 *Una successione* $(\mathbf{v}_n)$ *di* $\mathbf{R}^n$ *converge se e solo se è fondamentale.*

Per dire che in $\mathbf{R}^n$ le successioni convergenti sono tutte e sole le successioni fondamentali si dice che $\mathbf{R}^n$ è *completo*.

Una successione $(\mathbf{v}_n)$ si dice *limitata* la sua immagine è limitata, ossia quando esiste M tale che

$$\text{per ogni } n \text{ vale } \|\mathbf{v}_n\| < M\,.$$

Le disuguaglianze (3.3) mostrano che una successione è limitata se e solo se sono limitate le successioni delle sue componenti; e quindi anche in $\mathbf{R}^n$ vale il *Teorema di Bolzano-Weierstrass :*

Teorema 108 (di Bolzano-Weierstrass) *Ogni successione limitata ammette s.successioni convergenti.*

3.4 La norma euclidea

La norma di gran lunga più utile è la norma $\|\cdot\|_2$, perché essa ha una proprietà ben particolare, che ora illustriamo. Dai corsi di Geometria si sa che è possibile definire il *prodotto scalare* , detto anche *prodotto interno* , in $\mathbf{R}^n$ e che quando si conoscono le componenti di due vettori rispetto alla base canonica

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i\,, \qquad \mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{e}_i\,.$$

il prodotto scalare si calcola come[7]

$$\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n} \alpha_i \beta_i\,.$$

Si vede quindi che:

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}\cdot\mathbf{x}}\,. \tag{3.4}$$

Osservazione 109 Si osservi che il prodotto scalare di vettori associa a due vettori un numero (e non un vettore)! ∎

[7]talvolta invece che con vettori a componenti reali dovremo lavorare con vettori a componenti numeri complessi. In tal caso il prodotto scalare è $\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n} \alpha_i \bar{\beta}_i$ dove la barra indica il coniugato. Si noti che in questo modo $\mathbf{x}\cdot\mathbf{x}$ è sempre reale e positivo e la (3.4) vale anche per vettori a componenti numeri complessi.

Diciamo infine che un vettore che ha norma euclidea uguale ad 1 si chiama *versore*. I particolari versori **i**, **j**, **k** di $\mathbf{R}^3$ (o di $\mathbf{R}^2$ se non si considera **k**) si chiamano i *versori degli assi coordinati.*

Chiamiamo ora *ortogonali* due vettori che hanno prodotto scalare nullo e mostriamo che vale:

Teorema 110 (Teorema di Pitagora) *Se* $\mathbf{x}$ *e* $\mathbf{y}$ *sono due vettori di* $\mathbf{R}^n$*, si ha*

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 .$$

Dim. *Si ricordi dai corsi di geometria che il prodotto scalare gode della proprietà distributiva:*

$$\mathbf{x}\cdot(\mathbf{y}+\mathbf{z}) = \mathbf{x}\cdot\mathbf{y}+\mathbf{x}\cdot\mathbf{z} = (\mathbf{y}+\mathbf{z})\cdot\mathbf{x} .$$

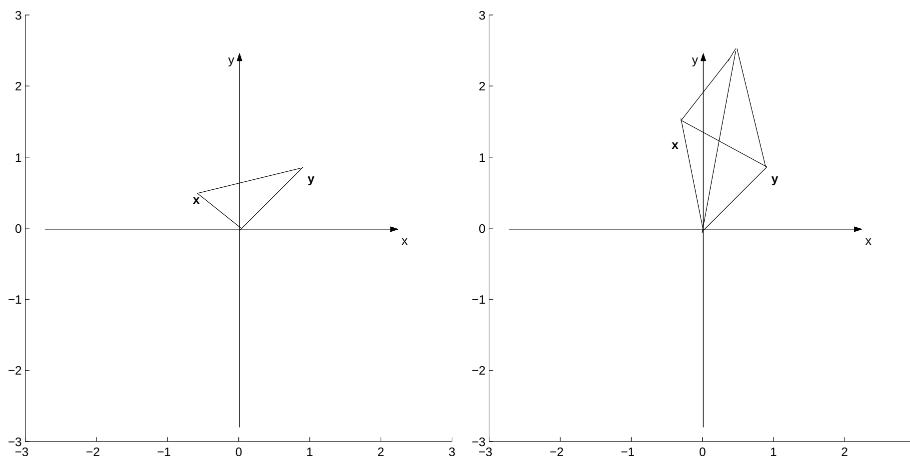
Dunque,

$$\begin{aligned}&\|\mathbf{x}+\mathbf{y}\|_2^2 = (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})\\ &= \mathbf{x}\cdot\mathbf{x} + (\mathbf{x}\cdot\mathbf{y}+\mathbf{y}\cdot\mathbf{x}) + \mathbf{y}\cdot\mathbf{y} = \|\mathbf{x}\|_2^2 + |\mathbf{y}\|_2^2\end{aligned}$$

perché ambedue gli addendi in parentesi sono nulli. ∎

Quando $\mathbf{x}$ e $\mathbf{y}$ sono ortogonali, il vettore $\mathbf{x}+\mathbf{y}$ è l'ipotenusa del triangolo rettangolo di cateti $\mathbf{x}$ e $\mathbf{y}$, si veda la figura 3.4. Dunque Il Teorema di Pitagora si interpetra dicendo che in $\mathbf{R}^2$ *il quadrato costruito sull'ipotenusa di un triangolo rettangolo ha per area la somma delle aree dei quadrati costruiti sui cateti.*

Figura 3.4: Teorema di Pitagora e identità del parallelogramma



Mostriamo ora una particolarissima proprietà della norma euclidea, che si chiama *identità del parallelogramma*.

Teorema 111 *Siano* $\mathbf{x}$ *e* $\mathbf{y}$ *due vettori di* $\mathbf{R}^n$. *Vale*

$$\|\mathbf{x}+\mathbf{y}\|^2+\|\mathbf{x}-\mathbf{y}\|^2=2\left[\|\mathbf{x}\|^2+\|\mathbf{y}\|^2\right] .$$

Dim. Usando la proprietà distributiva del prodotto scalare, calcoliamo

$$\begin{aligned}
&\|\mathbf{x}+\mathbf{y}\|^2+|\mathbf{x}-\mathbf{y}\|^2=(\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})+(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})\\
&=(\mathbf{x}\cdot\mathbf{x}+\mathbf{x}\cdot\mathbf{y}+\mathbf{y}\cdot\mathbf{x}+\mathbf{y}\cdot\mathbf{y})\\
&+(\mathbf{x}\cdot\mathbf{x}-\mathbf{x}\cdot\mathbf{y}-\mathbf{y}\cdot\mathbf{x}+\mathbf{y}\cdot\mathbf{y})=2\mathbf{x}\cdot\mathbf{x}+2\mathbf{y}\cdot\mathbf{y}=2\left[\|\mathbf{x}\|^2+\|\mathbf{y}\|^2\right] . \quad \blacksquare
\end{aligned}$$

La norma euclidea è l'unica norma che gode di questa proprietà. Per esercizio, si mostri che la proprietà del parallelogramma non vale per i vettori $\left[\begin{array}{cc} 0 & 1 \end{array}\right]'$ e $\left[\begin{array}{cc} 1 & 0 \end{array}\right]'$ di $\mathbf{R}^2$ con $\|\cdot\|_\infty$.

La figura 3.4 a destra mostra il significato dell'identità del parallelogramma in $\mathbf{R}^2$: $\mathbf{x}+\mathbf{y}$ e $\mathbf{x}-\mathbf{y}$ sono le diagonali del parallelogramma identificato dai due vettori $\mathbf{x}$ e $\mathbf{y}$ e quindi l'identità del parallelogramma è un'estensione del teorema di Pitagora: *in un parallelogramma, la somma delle aree dei quadrati costruiti sulle due diagonali è uguale alla somma delle aree dei quadrati costruiti sui quattro lati.*

Il prodotto scalare si definisce tra vettori liberi; la definizione si estende quindi al caso dei vettori applicati nel medesimo punto come si è detto al paragrafo 3.1.2.

Da ora in poi, se non si specifica esplicitamente il contrario, la norma in $\mathbf{R}^n$ sarà la norma $\|\cdot\|_2$, che indicheremo $\|\cdot\|$, sottintendendo l'indice. Useremo norme diverse per fare delle dimostrazioni se questo sarà conveniente. Infatti, il fatto che la relazione di convergenza non dipenda dalla particolare norma usata per verificarla può usarsi per semplificare alcune dimostrazioni.

3.4.1 R^2 e R^3 con la norma euclidea

Vogliamo ora esaminare più in dettaglio il caso di $\mathbf{R}^2$ e di $\mathbf{R}^3$ con norma euclidea. Diamo però la definizione seguente che vale anche in $\mathbf{R}^n$, dotato della norma euclidea e quindi del prodotto interno: due vettori $\mathbf{x}$ e $\mathbf{y}$ si dicono *ortogonali* quando hanno prodotto scalare nullo:

$$\mathbf{x}\perp\mathbf{y} \quad \text{quando } \mathbf{x}\cdot\mathbf{y}=0.$$

Si fissi ora il vettore $\mathbf{v}=a\mathbf{i}+b\mathbf{j}\neq\mathbf{0}$. Sono ad esso ortogonali i vettori $\mathbf{w}_1=b\mathbf{i}-a\mathbf{j}$ e $\mathbf{w}_2=-b\mathbf{i}+a\mathbf{j}$. Le matrici che trasformano la base canonica rispettivamente nella base $\{\mathbf{v},\mathbf{w}_1\}$ e $\{\mathbf{v},\mathbf{w}_2\}$ sono rispettivamente

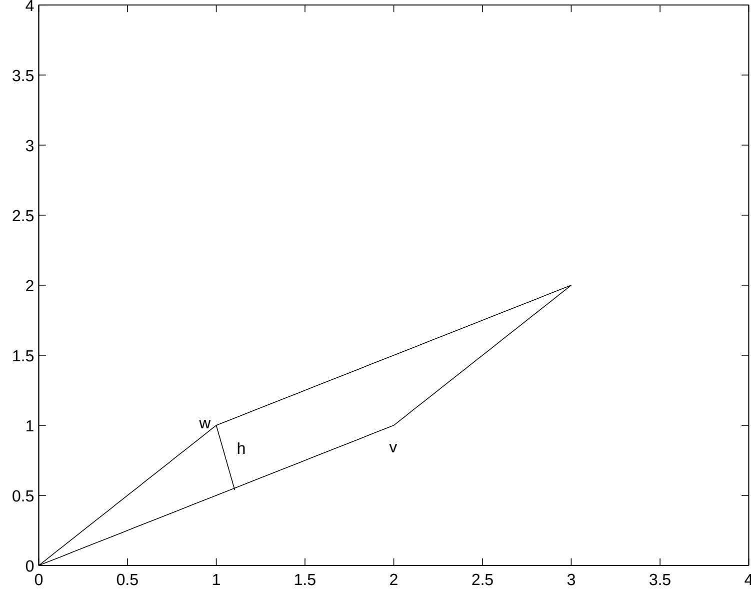
$$\left[\begin{array}{cc} a & b\\ b & -a \end{array}\right], \qquad \left[\begin{array}{cc} a & b\\ -b & a \end{array}\right].$$

La prima ha determinante negativo mentre la seconda ha determinante positivo. Dunque, la base $\{\mathbf{v},\mathbf{w}_2\}$ è orientata positivamente, ossia, come anche si dice, ha orientazione concorde con quella dello spazio. Per questa ragione, se non si specifica esplicitamente il contrario, come vettore ortogonale a $\mathbf{v}$ sceglieremo $-b\mathbf{i}+a\mathbf{j}$.

Vediamo ora di chiarire il significato geometrico dei determinanti di matrici 2×2 e 3×3.

Siano $\mathbf{v}$ e $\mathbf{w}$ due vettori, di componenti rispettivamente (a,b) e (c,d) rispetto alla base canonica di $\mathbf{R}^2$. Questi vettori identificano i due punti del piano $P\equiv(a,b)$ e $Q\equiv(c,d)$. Vogliamo calcolare l'area del parallelogramma in figura 3.5.

Figura 3.5: Calcolo dell'rea di un parallelogramma



Si sa che l'area è il prodotto della lunghezza di un lato per l'altezza ad esso relativa. Scegliendo come lato quello identificato dal vettore $\mathbf{v} = a\mathbf{i} + b\mathbf{j}$, vogliamo calcolare

$$\|\mathbf{v}\| \, \|\mathbf{h}\|$$

ove $\mathbf{h}$ è il vettore, applicato in Q, indicato in figura 3.5.

Dividendo per $\|\mathbf{v}\|$ non è restrittivo supporre $\|\mathbf{v}\| = 1$, ossia $a^2 + b^2 = 1$. In questo caso l'area è $\|\mathbf{h}\|$. L'altezza è il vettore

$$\mathbf{h} = \mathbf{w} - (\mathbf{w} \cdot \mathbf{v})\mathbf{v}\,.$$

Ricordando che $a^2 + b^2 = 1$, si calcola:

$$\begin{aligned}
&\|\mathbf{h}\|^2 = [\mathbf{w} - (\mathbf{w} \cdot \mathbf{v})\mathbf{v}] \cdot [\mathbf{w} - (\mathbf{w} \cdot \mathbf{v})\mathbf{v}] \\
&= [c - (ca + bd)a]^2 + [d - (ca + bd)b]^2 \\
&= c^2 - 2ac(ca + bd) + d^2 - 2bd(ca + bd) + (a^2 + b^2)(ca + bd)^2 \\
&= c^2(1 - a^2) + d^2(1 - b^2) - 2cabd = (cb)^2 + (da)^2 - 2cabd \\
&= (cb - ad)^2
\end{aligned}$$

e quindi l'area è

$$|cb - ad| = \left| \det \begin{bmatrix} a & c \\ b & d \end{bmatrix} \right| .$$

Si trova quindi un'interpretazione geometrica per il valore assoluto del determinante di una matrice 2×2: il numero $\left|\det \begin{bmatrix} \mathbf{v} & \mathbf{w} \end{bmatrix}\right|$ è l'area del parallelogramma identificato dai due

vettori $\mathbf{v}$ e $\mathbf{w}$ (applicati nell'origine). Il determinante stesso si interpretra anche come "area con segno" del parallelogramma identificato dai vettori che sono le colonne della matrice. Si può quindi concludere che valgono le affermazioni seguenti, per ogni coppia di vettori $\mathbf{v}$, $\mathbf{w}$ di $\mathbf{R}^2$ (in quest'ordine) e per la matrice

$$A = \begin{bmatrix} \mathbf{v} & \mathbf{w} \end{bmatrix} :$$

- Condizione necessaria e sufficiente perché i due vettori siano colineari è che il determinante di A sia nullo;
- se il determinante è non nullo, i due vettori $\mathbf{v}$ e $\mathbf{w}$ (in quest'ordine) sono una base ordinata di $\mathbf{R}^2$; il parallelogramma che essi identificano ha "area con segno" che è positiva se e solo se essi, presi nell'ordine dato, sono una base orientata positivamente;
- la matrice A subordina una trasformazione lineare in $\mathbf{R}^2$. Siano $\mathbf{r}$ e $\mathbf{r}'$ due vettori (applicati nell'origine) e $\tilde{\mathbf{r}} = A\mathbf{r}$, $\tilde{\mathbf{r}}' = A\mathbf{r}'$ le loro immagini mediante la matrice A. Sia R il parallelogramma identificato dai vettori $\mathbf{r}$ e $\mathbf{r}'$ e sia $\tilde{R}$ il parallelogramma identificato dalle loro immagini. L'area del parallelogramma $\tilde{R}$ è il prodotto di $|\det A|$ per l'area del parallelogramma R:

$$(\text{area di } \tilde{R}) = |\det A| \cdot (\text{area di } R). \tag{3.5}$$

Risultati del tutto analoghi valgono anche in $\mathbf{R}^3$. Dati tre vettori $\mathbf{u}$, $\mathbf{v}$ e $\mathbf{w}$, in quest'ordine, si costruisce la matrice

$$A = \begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{bmatrix} .$$

Il valore assoluto del suo determinante è il volume del parallelepipedo identificato dai tre vettori (applicati nell'origine). La matrice A identifica una trasformazione lineare. Tale trasformazione applicata ai punti di un parallelepipedo lo trasforma in un altro, il cui volume differisce da quello del primo per il fattore $|\det A|$.

Il numero $\det A$ si interpreta come "volume con segno".

3.5 Il prodotto vettoriale

A differenza delle operazioni tra vettori introdotte fino ad ora, che valgono in $\mathbf{R}^n$ per ogni n, l'operazione di prodotto vettoriale è specifica di $\mathbf{R}^3$. Essa si definisce ponendo:

$$\begin{array}{lll} \mathbf{i} \wedge \mathbf{i} = 0\,, & \mathbf{j} \wedge \mathbf{j} = 0\,, & \mathbf{k} \wedge \mathbf{k} = 0\,, \\ \mathbf{i} \wedge \mathbf{j} = \mathbf{k}\,, & \mathbf{j} \wedge \mathbf{k} = \mathbf{i}\,, & \mathbf{k} \wedge \mathbf{i} = \mathbf{j}\,, \\ \mathbf{i} \wedge \mathbf{j} = -\mathbf{j} \wedge \mathbf{i}\,, & \mathbf{j} \wedge \mathbf{k} = -\mathbf{k} \wedge \mathbf{j}\,, & \mathbf{k} \wedge \mathbf{i} = -\mathbf{i} \wedge \mathbf{k}\,. \end{array}$$

Completiamo ora la definizione di prodotto vettoriale ponendo

$$\mathbf{v} \wedge \mathbf{w} = (a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) \wedge (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = (bz - cy)\mathbf{i} + (cx - az)\mathbf{j} + (ay - bx)\mathbf{k}$$

formalmente ottenuta distribuendo le somme sui prodotti e facendo uso delle regole per i prodotti vettoriali degli elementi della base. Di conseguenza, si verificano le regole seguenti:

1. $\mathbf{v} \wedge \mathbf{w} = -\mathbf{w} \wedge \mathbf{v}$;
2. $\mathbf{v} \wedge \mathbf{w}$ è ortogonale sia a $\mathbf{v}$ che a $\mathbf{w}$;

3. vale:
$$(\mathbf{r} \wedge \mathbf{v}) \cdot \mathbf{w} = \det \begin{bmatrix} \mathbf{r} & \mathbf{v} & \mathbf{w} \end{bmatrix} .$$
Quindi, il valore assoluto $|(\mathbf{r} \wedge \mathbf{v}) \cdot \mathbf{w}|$ è il volume del parallelepipedo identificato dai tre vettori (pensati applicati nell'origine);

4. In particolare,
$$\det \begin{bmatrix} \mathbf{r} & \mathbf{v} & \mathbf{r} \wedge \mathbf{v} \end{bmatrix} = (\mathbf{r} \wedge \mathbf{v}) \cdot (\mathbf{r} \wedge \mathbf{v}) .$$
Questo numero è zero se i vettori $\mathbf{r}$ e $\mathbf{v}$ sono colineari. Altrimenti è positivo. Dunque, i tre vettori $\mathbf{r}$, $\mathbf{v}$, $\mathbf{r} \wedge \mathbf{v}$ (presi in quest'ordine e con $\mathbf{r}$ e $\mathbf{v}$ non colineari) subordinano in $\mathbf{R}^3$ l'orientazione positiva.

Si noti che:

- Il prodotto vettoriale si definisce per vettori liberi; la definizione si estende quindi al caso dei vettori applicati nel medesimo punto come si è detto al paragrafo 3.1.2.
- Il prodotto vettoriale di due vettori è un vettore. Per contrasto, il prodotto scalare di due vettori è un numero.
- Il prodotto vettoriale è nullo se e solo se i due vettori sono colineari. Per contrasto, il prodotto scalare è nullo se e solo se i due vettori sono ortogonali.

Infine, si noti che le regole per il calcolo del prodotto vettoriale sono definite in modo da "mimare" quelle per il calcolo dei determinanti.

Osservazione 112 Il prodotto vettoriale può definirsi in particolare per vettori complanari, per esempio per vettori del piano $z = 0$. In tal caso il prodotto vettoriale è un vettore "verticale", ossia parallelo all'asse z. ∎

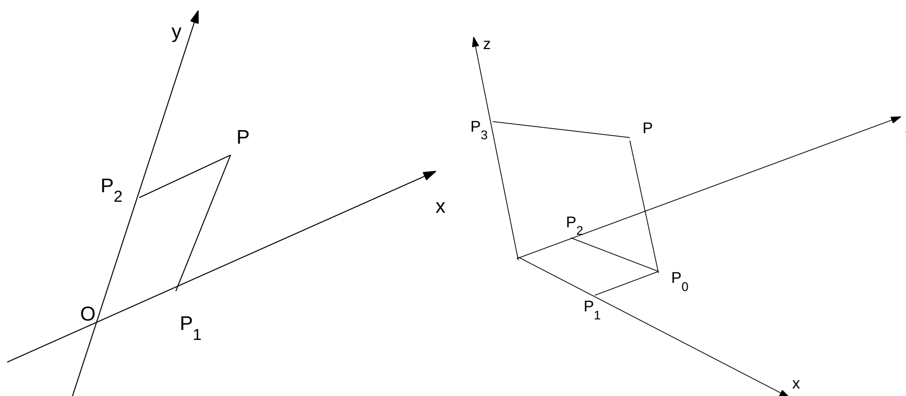
Osservazione sulla notazione La notazione col punto, $\mathbf{v} \cdot \mathbf{w}$, per indicare il prodotto scalare è oggi universalmente usata[8]. Invece, la notazione per il prodotto vettoriale non è così uniforme. La notazione $\mathbf{v} \wedge \mathbf{w}$ si trova principalmente in testi europei, mentre in testi americani (ed anche inglesi) il prodotto vettoriale è indicato $\mathbf{v} \times \mathbf{w}$.

3.6 Coordinate curvilinee nel piano e nello spazio

Il modo più comune per rappresentare i punti del piano, o dello spazio, usa le coordinate cartesiane ortogonali. Però, punti del piano e dello spazio possono rappresentarsi, oltre che in coordinate cartesiane ortogonali mediante *coordinate cartesiane* *oblique* o anche con altri "sistemi di coordinate" che generalmente costruiscono corrispondenze biunivoche tra i punti (del piano o dello spazio) (o talvolta di opportuni loro s.insiemi) ed opportuni s.insiemi di $\mathbf{R}^2$ oppure di $\mathbf{R}^3$. Per ragioni che vedremo, si parla in tal caso di *coordinate curvilinee*. Noi esamineremo prima le coordinate cartesiane oblique (nel piano. La semplice estensione

[8]in libri molto vecchi e assai raramente in testi recenti si trova usata la croce per il prodotto scalare, $\mathbf{v} \times \mathbf{w}$.

Figura 3.6: Coordinate oblique



allo spazio è lasciata al lettore). Poi studieremo alcuni casi particolari di coordinate curvilinee: vedremo l'uso delle *coordinate polari* e delle *coordinate ellittiche* per rappresentare i punti del piano cartesiano e l'uso delle *coordinate cilindriche*, *sferiche* ed *ellittiche* per rappresentare i punti dello spazio, che supporremo dotato di un sistema di coordinate cartesiane ortogonali.

Coordinate cartesiane oblique nel piano (e nello spazio). Fissata l'origine O, tracciamo per essa due rette non coincidenti (tre rette non complanari nello spazio). Queste rette si chiamano *assi cartesiani obliqui*.

Si decida quale è la prima retta, asse delle ascisse, la seconda, asse delle ordinate, un verso positivo su di esse e un'unità di misura (che potrebbe anche essere diversa. Noi assumeremo che sia la medesima). Da un punto P facciamo uscire due rette, parallele agli assi delle ordinate e delle ascisse. La retta parallela all'asse delle ordinate interseca l'asse delle ascisse in un punto P_1 la cui distanza distanza (con segno) da O si chiama l'*ascissa* di P. In modo analogo si definisce l'*ordinata* di P. La coppia ordinata dell'ascissa e dell'ordinata rappresenta univocamente il punto P. Si veda la figura 3.6.

Un problema importante è di passare da un sistema di coordinate ad un altro. Limitiamoci a studiare il caso di due sistemi di coordinate, uno un sistema di coordinate cartesiane ortogonali del piano ed uno un sistema di coordinate oblique. Provvisoriamente indichiamo con lettere greche le coordinate oblique: ξ è l'asse delle ascisse (oblique) ed η quello delle ordinate.

Indichiamo con $\mathbf{i}$ e $\mathbf{j}$ i versori degli assi cartesiani ortogonali e siano

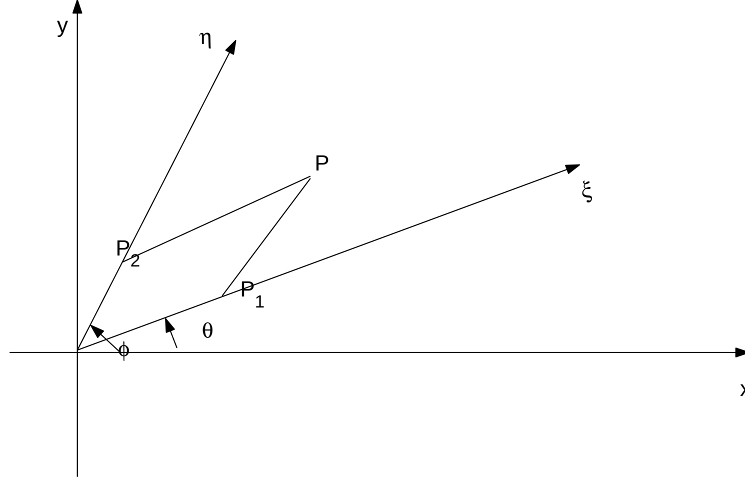
$$\mathbf{v}_1 = \mathbf{i}\cos\theta + \mathbf{j}\sin\theta\,, \qquad \mathbf{v}_2 = \mathbf{i}\cos\phi + \mathbf{j}\sin\phi$$

i versori degli assi obliqui, si veda la figura 3.7.

Si conoscano le coordinate cartesiane ortogonali (x, y) del punto P. Le coordinate oblique di P sono le distanze dall'origine dei vettori dei punti P_1 e P_2, che avranno forma

$$\xi\mathbf{v}_1\,, \qquad \eta\mathbf{v}_2$$

Figura 3.7: Trasformazione a coordinate oblique



per certe scelte dei parametri ξ e η. Poiché $\mathbf{v}_1$ e $\mathbf{v}_2$ sono versori, le coordinate oblique di P sono proprio i numeri ξ e η, che ora vogliamo calcolare. Ciò si fa notando che

$$x\mathbf{i} + y\mathbf{j} = \xi\mathbf{v}_1 + \eta\mathbf{v}_2 = \xi(\mathbf{i}\cos\theta + \mathbf{j}\sin\theta) + \eta(\mathbf{i}\cos\phi + \mathbf{j}\sin\phi)\,.$$

Uguagliando le componenti si ottiene immediatamente

$$\begin{cases} \xi = \frac{y\cos\phi - x\sin\phi}{\sin(\theta-\phi)}\,, \\ \eta = \frac{x\sin\theta - y\cos\theta}{\sin(\theta-\phi)}\,, \end{cases} \qquad \begin{cases} x = \xi\cos\theta + \eta\cos\phi \\ y = \xi\sin\theta + \eta\sin\phi\,. \end{cases}$$

Si noti che il denominatore è non nullo perchè gli assi obliqui non coincidono.

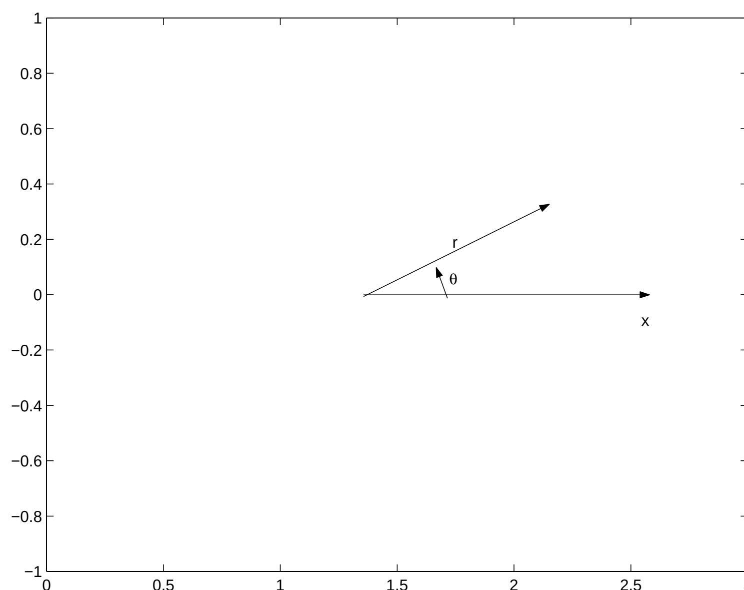
Si osservi un caso particolare: supponiamo che $\phi - \theta = \pi/2$. In questo caso gli "assi obliqui" sono tra loro perpendicolari e si vuol rappresentare il medesimo punto P rispetto a due sistemi di assi cartesiani ortogonali ruotati l'uno rispetto all'altro. Precisamente, il secondo sistema è ottenuto ruotando il primo dell'angolo θ (in senso positivo o negativo). Essendo in questo caso particolare $\phi = \theta + \pi/2$, le coordinate (ξ, η) sono date da

$$\begin{cases} \xi = x\cos\theta + y\sin\theta \\ \eta = -x\sin\theta + y\cos\theta\,, \end{cases} \qquad \begin{cases} x = \xi\cos\theta - \eta\sin\theta \\ y = \xi\sin\theta + \eta\cos\theta\,. \end{cases}$$

Coordinate polari nel piano. Sia P il punto da rappresentare. Si rappresenta P mediante la sua distanza da O e mediante l'angolo θ tra la retta r che esce dall'origine O e punta verso P e il semiasse positivo delle ascisse. L'angolo si sceglie col segno in questo modo: si orienta la retta r da O verso P; si riporta il semiasse $x > 0$ sulla semiretta r, ruotando dell'angolo minore. L'angolo θ così ottenuto si intende positivo se la rotazione è antioraria (si confronti con la regola d'Ampère).

In questo modo si stabilisce una corrispondenza biunivoca tra i punti del piano (x, y), escluso O, e le coppie di numeri (ρ, θ) con $\rho > 0$ e $0 \le \theta < 2\pi$.

Figura 3.8: coordinate polari



Dunque, ogni punto P del piano (x, y), escluso O, viene identificato dalla coppia dei numeri ρ e θ, che si chiamano le *coordinate polari* di P.

L'origine invece è identificata da $(0, \theta)$ per ogni θ. Si veda la figura 3.8. Il numero ρ si chiama il definmodulo e θ si chiama l'*argomento*}*o* *anomalia* }o *anomalia* di P.

La relazione tra le coordinate cartesiane e le coordinate polari è data da:

$$x = \rho \cos \theta , \qquad y = \rho \sin \theta .$$

nel contesto delle coordinate polari, il semiasse positivo delle ascisse si chiama anche *asse polare* e il suo estremo, ossia l'origine, si chiama anche *polo*.

Osservazione 113 *Va notato che la corrispondenza* $(\rho, \theta) \mapsto (x, y)$ *è suriettiva ma non iniettiva; dunque non invertibile. Si trova una corrispondenza biunivoca di* $\mathbf{R}^2 \setminus \{0\}$ *se* $(\rho, \theta) \in (0, +\infty) \times [0, 2\pi)$.

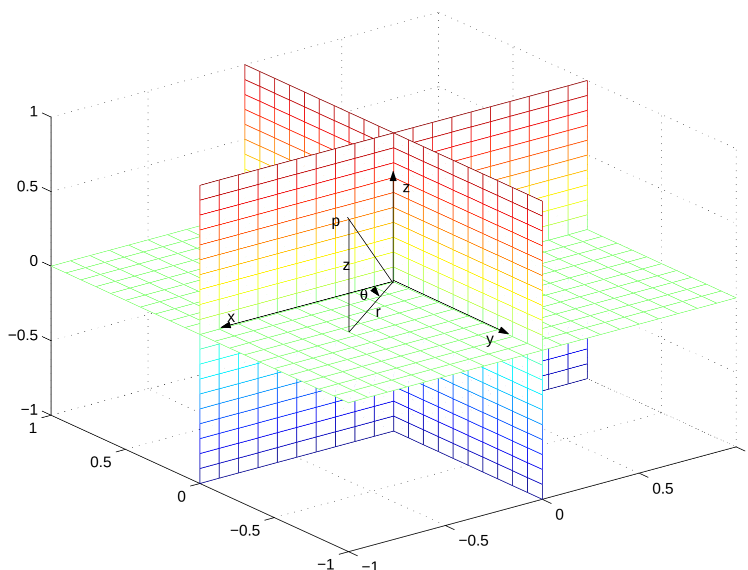
Coordinate polari ellittiche piano. Siano assegnati due numeri positivi a e b. Le *coordinate polari ellittiche* nel piano si ottengono rappresentando i punti (x, y) mediante i numeri (ρ, θ) tali che

$$x = \rho a \cos \theta , \qquad y = \rho b \sin \theta .$$

Le curve coordinate delle coordinate polari od ellittiche. Torniamo alle relazioni

$$x = \rho a \cos \theta , \qquad y = \rho b \sin \theta .$$

Figura 3.9: coordinate cilindriche



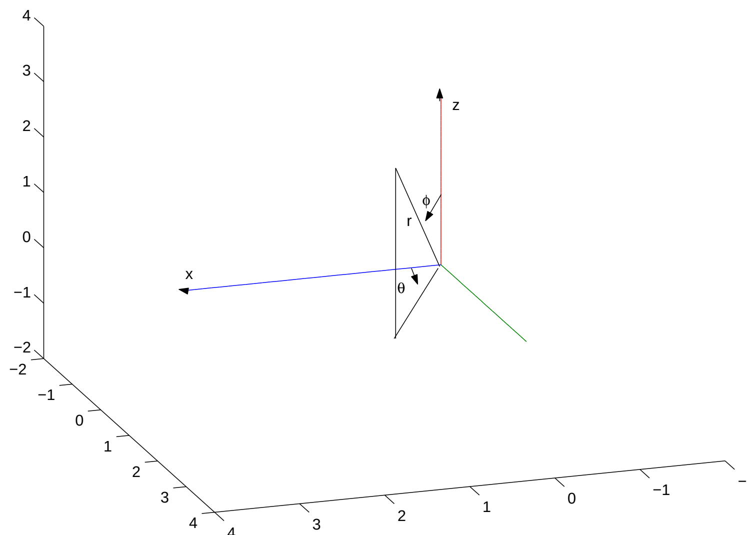
Queste relazioni identificano un punto (x, y) del piano per ogni scelta di (ρ, θ). Si chiamano *curve coordinate* quelle ottenute da queste espressioni per θ fissato al variare di r (rette per l'origine) e per r fissato al variare di θ (ellissi; nel caso particolare delle coordinate polari si hanno circonferenze).

Coordinate cilindriche nello spazio. Sia $P \equiv (x, y, z)$ un punto dello spazio, riferito ad un sistema di coordinate cartesiane ortogonali. Il punto $Q \equiv (x, y, 0)$ si chiama la *proiezione ortogonale* di P sul piano $z = 0$. Il punto Q si può rappresentare mediante le sue coordinate polari (ρ, θ) e quindi P viene ad essere rappresentato mediante (ρ, θ, z). Quando si fa uso di questa rappresentazione si dice che si rappresenta lo spazio in *coordinate cilindriche*, si veda la figura 3.9.

Se invece delle coordinate polari, sul piano si usano le coordinate ellittiche, le corrispondenti coordinate nello spazio si chiamano *cilindriche ellittiche*.

Le superfici coordinate delle coordinate cilindriche. Le curve coordinate sono quelle curve che si ottengono tenendo fissi i valori di due parametri e facendo variare il terzo. Sono quindi rette per l'origine, circonferenze (o ellissi) e rette verticali. Però nello spazio si possono anche definire le *superfici coordinate* , ottenute tenendo fisso un parametro e facendo variare gli altri due. Quindi, nel caso delle coordinate cilindriche ellittiche, le superfici coordinate sono cilindri ellittici di asse parallelo all'asse z (ottenuti tenendo fisso il valore di ρ); piani per l'asse z (ottenuti tenendo fisso il valore di θ); piani perpendicolari all'asse z (ottenuti tenendo fisso il valore di z).

Figura 3.10: coordinate sferiche



Coordinate sferiche ed ellittiche nello spazio. Le coordinate sferiche nello spazio sono l'analogo delle coordinate polari nel piano. Per rappresentare un punto $P(x, y, z)$, si costruisce la retta congiungente O con P. Si rappresenta P mediante (r, θ, ϕ) dove r è la distanza di P da O e ancora si chiama l' *argomento* di P; θ è l'argomento della proiezione Q di P sul piano $z = 0$, ϕ è l'angolo tra il versore ν che sulla retta da O a P punta verso P ed il versore $\mathbf{k}$. L'ampiezza di quest'angolo si intende compresa tra 0 e π.

Dunque, si veda la figura 3.10, P si rappresenta anche con la terna (r, θ, ϕ) con

$$r \geq 0\,, \qquad 0 \leq \theta < 2\pi\,, \qquad 0 \leq \phi \leq \pi\,.$$

Questi numeri si chiamano le *coordinate sferiche* di P. La relazione tra le coordinate cartesiane e le coordinate sferiche è la seguente:

$$x = r\cos\theta\sin\phi\,, \qquad y = r\sin\theta\sin\phi\,, \qquad z = r\cos\phi\,.$$

Il numero r si chiama ancora *modulo*, il numero θ si chiama *longitudine* mentre il numero ϕ si chiama *colatitudine*.

Siano dati ora tre numeri positivi a, b e c. Le coordinate ellittiche di un punto nello spazio sono le coordinate (ρ, θ, ϕ) che si ottengono imponendo

$$x = ra\cos\theta\sin\phi\,, \qquad y = rb\sin\theta\sin\phi\,, \qquad z = rc\cos\phi\,.$$

Le superfici coordinate delle coordinate sferiche ed ellittiche. Nel caso delle coordinate sferiche, le superfici coordinate sono sfere di centro l'origine (ottenute

tenendo fisso il valore di r); piani per l'asse z (ottenuti fissando il valore di θ; (semi)coni circolari di asse sull'asse z (ottenuti tenendo fisso il valore di ϕ).

Per esercizio, se ne identifichino le curve coordinate e si identifichino anche le curve e le superfici coordinate delle coordinate ellittiche.

3.7 Funzioni da R in R^n

Studiamo ora le proprietà di limite e continuità delle funzioni definite su sottoinsiemi di $\mathbf{R}$ ed a valori in $\mathbf{R}^n$.

Sia $\mathbf{x} \in \mathbf{R}^n$ un vettore e siano x_i le sue componenti. E' utile tener presente le disuguaglianze seguenti (si vedano le (3.3)):

- Per ogni i vale $|x_i| \leq \|\mathbf{x}\|$;
- Esiste un numero $M = M_{n,p}$ tale che $\|\mathbf{x}\|_p \leq M \sum_{i=1}^n |x_i|$.

Le funzioni da $\mathbf{R}$ in $\mathbf{R}$ sono state studiate nel corso di Analisi Matematica 1. E' facile adattare gli argomenti visti nel corso di Analisi Matematica 1 al caso di funzioni a valori vettoriali. Per l'uso che a noi servirà, consideriamo funzioni definite su un intervallo I, aperto o meno, limitato o meno, a valori in $\mathbf{R}^n$. Dunque, si specifica la funzione assegnando ad ogni valore $t \in I$ un vettore $\mathbf{x}(t) \in \mathbf{R}^n$, ossia assegnando le sue n componenti. Si costruisce così una funzione

$$t \to \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} = x_1(t)\mathbf{e}_1 + x_2(t)\mathbf{e}_2 + \cdots + x_n(t)\mathbf{e}_n \,.$$

Nel caso di $\mathbf{R}^2$ o di $\mathbf{R}^3$ scriveremo anche

$$\mathbf{x}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \,.$$

Notiamo che le componenti di $\mathbf{x}(t)$ sono funzioni da $\mathbf{R}$ in $\mathbf{R}$ e quindi le nozioni apprese nel corso di Analisi 1 possono essere applicate a ciascuna componente. Ora definiamo:

- **Limiti e continuità**

Si fissi un punto t_0. Il punto t_0 può essere un punto di I o anche un estremo di I che non gli appartiene.

Si dice che $\boxed{\lim_{t \to t_0} \mathbf{x}(t) = \mathbf{l}}$ quando assegnata una qualsiasi palla $B(\mathbf{l}, \epsilon)$ centrata in $\mathbf{l}$ esiste un intorno $I_\epsilon(t_0)$ tale che se $t \in I_\epsilon \cap I$, $\boxed{t \neq t_0}$, allora $\mathbf{x}(t) \in B(\mathbf{l}, \epsilon)$. In simboli,

$$\forall \epsilon \quad \exists \delta \mid t \in I \text{ e } 0 < |t - t_0| < \delta \Rightarrow \|\mathbf{x}(t) - \mathbf{l}\| < \epsilon \,.$$

Se $t \in I$ e inoltre

$$\lim_{t \to t_0} \mathbf{x}(t) = \mathbf{x}(t_0)$$

si dice che la funzione è $\boxed{\textit{continua}}$ in t_0.

Si lascia per esercizio di definire i limiti per $t \to +\infty$ e per $t \to -\infty$, sulla falsariga della definizione del limite di successioni, vista al paragrafo 3.3. Si noti però che non è possibile definire limiti uguali a $\pm\infty$. Se $\mathbf{x}(t)$ non rimane limitata allora possiamo definire solamente $\lim \|\mathbf{x}(t)\| = +\infty$ e si ricade in un caso già trattato nel corso di Analisi Matematica 1 perché la funzione $t \to \|\mathbf{x}(t)\|$ è una funzione da $\mathbf{R}$ in sé.

Siano $x_i(t)$ le componenti di $\mathbf{x}(t)$ ed l_i quelle di $\mathbf{l}$. Il teorema seguente è analogo al Teorema **??**.

Teorema 114 *Vale* $\lim \mathbf{x}(t) = \mathbf{l}$ *se e solo se per ogni* i *vale* $\lim x_i(t) = l_i$.

Una funzione $t \to \mathbf{x}(t)$ *da* $\mathbf{R}$ *in* $\mathbf{R}^n$ *è continua se e solo se ciascuna sua componente è una funzione continua da* $\mathbf{R}$ *in* $\mathbf{R}$.

Noto ciò, è immediato dedurre il risultato seguente:

Teorema 115 *Siano* $t \to \mathbf{x}(t)$ *e* $t \to \mathbf{y}(t)$ *due funzioni da* $\mathbf{R}$ *in* $\mathbf{R}^n$ *definite sul medesimo intervallo* I *ed ambedue continue e sia* $t \to k(t)$ *una funzione da* $\mathbf{R}$ *in* $\mathbf{R}$ *definita su* I *e continua. Le funzioni*

$$t \to k(t)\mathbf{x}(t)\,, \qquad t \to \mathbf{x}(t) + \mathbf{y}(t)\,, \qquad t \to \mathbf{x}(t) \cdot \mathbf{y}(t)$$

(il punto indica il prodotto scalare calcolato per ogni valore di t*) sono continue.*

Se $n = 3$*, anche la funzione* $t \to \mathbf{x}(t) \wedge \mathbf{y}(t)$ *è continua.*

Si lascia per esercizio di enunciare i teoremi corrispondenti per i limiti.

- **i simboli di Landau**

  Si dice che $\mathbf{x}(t)$ è un *infinitesimo* per $t \to t_0$ quando

$$\lim_{t \to t_0} \mathbf{x}(t) = 0\,.$$

Siano $\mathbf{x}(t)$ e $\mathbf{y}(t)$ due funzioni definite sul medesimo intervallo I e se ne vogliano studiare le relazioni per $t \to t_0$. Diciamo che:

i) $\mathbf{x} = \mathbf{O}(\mathbf{y})$ quando esistono un numero M ed un intorno J di t_0 tali che:

$$t \in I \cap J \;\Rightarrow\; \|\mathbf{x}(t)\| \le M\|\mathbf{y}(t)\|\,.$$

ii) $\mathbf{x} = \mathbf{o}(\mathbf{y})$ quando $\|y(t)\| \ne 0$ per $t \ne t_0$ in un intorno di t_0 e inoltre

$$\lim_{t \to t_0} \frac{\|\mathbf{x}(t)\|}{\|\mathbf{y}(t)\|} = 0\,.$$

In queste definizioni le due funzioni $\mathbf{x}(t)$ e $\mathbf{y}(t)$ potrebbero avere valori in spazi di dimensione diversa. In particolare, $\mathbf{y}(t)$ potrebbe essere una funzione a valori in $\mathbf{R}$. Se $y(t)$ prende valori in $\mathbf{R}$ ed è costantemente uguale ad 1, la condizione $\mathbf{x} = \mathbf{o}(1)$ vul dire che $\mathbf{x}(t)$ è un infinitesimo (sottinteso, per $t \to t_0$).

- **La derivabilità**

Sia t_0 punto interno di I. Si dice che un vettore $\mathbf{l}$ è la *derivata* di $\mathbf{x}(t)$ quando

$$\mathbf{l} = \lim_{t\to t_0} \frac{\mathbf{x}(t) - \mathbf{x}(t_0)}{t - t_0}\,.$$

La derivata si indica con uno dei soliti simboli,

$$\frac{\mathrm{d}\mathbf{x}(t_0)}{\mathrm{d}t}\,, \qquad \mathbf{x}'(t_0)\,, \qquad \dot{\mathbf{x}}(t_0)\,, \qquad D\mathbf{x}(t_0)\,, \qquad D_{t_0}\mathbf{x} \quad \text{ecc.}$$

Se t indica il tempo e se $\mathbf{x}(t)$ indica la posizione di un punto all'istante t, allora il quoziente

$$\frac{\mathbf{x}(t) - \mathbf{x}(t_0)}{t - t_0}$$

indica la velocità media del punto, nell'intervallo di tempo (t_0, t); e quindi la derivata si interpreta come *velocità* del punto all'istante t.

Usando le disuguaglianzein (3.3), si vede che

Teorema 116 *La funzione $\mathbf{x}(t)$ è derivabile in t_0 se e solo se ciascuna sua componente è derivabile in $\mathbf{t}_0$ e inoltre*

$$\mathbf{x}'(t_0) = \begin{bmatrix} x_1'(t_0) \\ x_2'(t_0) \\ \vdots \\ x_n'(t_0) \end{bmatrix}.$$

Dunque, una funzione derivabile in un punto t_0 è ivi continua.

In modo del tutto analogo si definiscono le derivate direzionali.

Le usuali regole di calcolo delle derivate si possono applicare alle singole componenti del vettore $\mathbf{x}(t)$ e quindi, per esempio, vale ancora la proprietà di *linearità* *della derivata:*

$$D_{t_0}\{\mathbf{x}(t) + \mathbf{y}(t)\} = \mathbf{x}'(t_0) + \mathbf{y}'(t_0)\,.$$

Teorema 117 *Siano $\mathbf{x}(t)$ ed $\mathbf{y}(t)$ derivabili in t_0. I prodotti scalare e vettoriale*[9] *sono derivabili in t_0 e valgono le uguaglianze*

$$\frac{d}{dt}\left[\mathbf{x}(t)\cdot\mathbf{y}(t)\right] = \mathbf{x}'(t)\cdot\mathbf{y}(t) + \mathbf{x}(t)\cdot\mathbf{y}'(t)\,,$$
$$\frac{d}{dt}\left[\mathbf{x}(t)\wedge\mathbf{y}(t)\right] = \mathbf{x}'(t)\wedge\mathbf{y}(t) + \mathbf{x}(t)\wedge\mathbf{y}'(t)\,.$$

Dim. Proviamo l'asserto per il prodotto vettoriale.

Bisogna calcolare

$$\lim_{h\to 0} \frac{\mathbf{x}(t+h)\wedge\mathbf{y}(t+h) - \mathbf{x}(t)\wedge\mathbf{y}(t)}{h}\,.$$

Aggiungendo e sottraendo al numeratore $\mathbf{x}(t)\wedge\mathbf{y}(t+h)$ si vede che la derivata è uguale a

$$\lim_{h\to 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h}\wedge\mathbf{y}(t+h) + \lim_{h\to 0}\mathbf{x}(t)\wedge\frac{\mathbf{y}(t+h) - \mathbf{y}(t)}{h} = \mathbf{x}'(t)\wedge\mathbf{y}(t) + \mathbf{x}(t)\wedge\mathbf{y}'(t)$$

[9] ricordiamo, definito solo in $\mathbf{R}^3$.

perchè

$$\lim_{h\to 0} y(t+h) = y(t)$$

perché la funzione $\mathbf{y}(t)$, essendo derivabile, è anche continua

La dimostrazione per il prodotto scalare è simile. ∎

Osservazione 118 Nella regola per la derivata del prodotto vettoriale i fattori non possono scambiarsi; invece la derivata del prodotto scalare non dipende dall'ordine dei fattori. ∎

In particolare:

Teorema 119 *Sia* $\mathbf{x}(t)$ *una funzione derivabile a valori in* $\mathbf{R}^n$ *e tale che* $\|\mathbf{x}(t)\| \equiv 1$*. Allora,* $\mathbf{x}'(t) \perp \mathbf{x}(t)$ *.*

Dim. Infatti, derivando i due membri dell'uguaglianza

$$1 = \|\mathbf{x}(t)\|^2 = \mathbf{x}(t)\cdot\mathbf{x}(t)$$

si trova

$$2\mathbf{x}'(t)\cdot\mathbf{x}(t) \equiv 0$$

ossia

$$\mathbf{x}'(t) \perp \mathbf{x}(t)\,. \quad ∎$$

Ovviamente, se $\mathbf{x}(t)$ è derivabile in t_0, vale la *prima formula degli* *incrementi finiti*

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \mathbf{x}'(t_0)(t-t_0) + \mathbf{o}(t-t_0)\,.$$

Osservazione 120 Va esplicitamente notato che la seconda formula degli incrementi finiti, ossia il Teorema di Lagrange, NON vale. Per rendersi conto di ciò, consideriamo la funzione $\mathbf{x}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$. Il Teorema di Lagrange può applicarsi alle due componenti separatamente, ottenendo

$$x(t') - x(t'') = x'(c_1)(t'-t'')\,, \qquad y(t') - y(t'') = y'(c_2)(t'-t'')$$

e generalmente $c_1 \neq c_2$. ∎

Se le singole componenti di $\mathbf{x}(t)$ sono ciascuna derivabile 2 volte, potremo introdurre le derivate seconde e, in generale, le derivate k-me in t_0. Se ciascuna componente ammette k derivate in t_0 si può anche scrivere la *formula di* *Taylor* con resto in forma di Peano,

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \sum_{j=1}^{k} \frac{\mathbf{x}^{(j)}(t_0)}{j!}(t-t_0)^j + \mathbf{o}(t-t_0)^k\,.$$

- **Integrale**

Se $t \to \mathbf{x}(t)$ è definita su $[a,b]$ e ciascuna sua componente è integrabile, si definisce $\int_a^b \mathbf{x}(t)\,\mathrm{d}t$ come quel vettore che ha per componenti i numeri $\int_a^b x_i(t)\,\mathrm{d}t$.

Capitolo 4

Funzioni da $\mathbf{R}^n$ in $\mathbf{R}^m$

Studiamo ora le funzioni di più di una variabile; ossia funzioni

$$\mathbf{F}(\mathbf{x})$$

a valori in $\mathbf{R}^m$, della variabile $\mathbf{x} \in \mathbf{R}^n$, in generale con $n \neq m$. Ovviamente, le prime proprietà da studiare saranno le proprietà relative ai limiti ed alla continuità. Come vedremo, non si incontrano situazioni sostanzialmente diverse da quelle delle funzioni da $\mathbf{R}$ in $\mathbf{R}$ (però le proprietà relative agli estremi saranno tipiche delle funzioni a valori reali). Differenze profonde si troveranno invece nello studio della differenziabilità.

4.1 Limiti e continuità

Ricordiamo che col simbolo $\|\mathbf{r}\|$ intendiamo la norma di $\mathbf{r}$, senza esplicitamente indicare la dimensione del vettore. Per esempio,

$$\begin{aligned}
&\text{se } \mathbf{r} = x \in \mathbf{R} \text{ allora } \|\mathbf{r}\| = \sqrt{x^2} = |x|;\\
&\text{se } \mathbf{r} = (x, y) \text{ allora } \|\mathbf{r}\| = \sqrt{x^2 + y^2};\\
&\text{se } \mathbf{r} = (x, y, z) \text{ allora } \|\mathbf{r}\| = \sqrt{x^2 + y^2 + z^2}.
\end{aligned}$$

Se vorremo specificare che la norma è quella di $\mathbf{R}^n$ allora scriveremo

$$\|\mathbf{r}\|_{\mathbf{R}^n}.$$

Sia $\mathbf{F}(\mathbf{r})$ una funzione definita su un insieme Ω di $\mathbf{R}^n$, a valori in $\mathbf{R}^m$ e sia $\mathbf{r_0}$ un punto di accumulazione di Ω.

Diremo che

$$\lim_{\mathbf{r}\to\mathbf{r_0}} \mathbf{F}(\mathbf{r}) = \mathbf{L}$$

Se per **ogni** $\epsilon > 0$ **esiste** un intorno $I_\epsilon(\mathbf{r_0})$ del punto $\mathbf{r_0}$ tale che

per **ogni** $\mathbf{r} \in \Omega \cap I_\epsilon(\mathbf{r_0})$, $\boxed{\mathbf{r} \neq \mathbf{r_0}}$ si ha

$$\|\mathbf{F}(\mathbf{r}) - \mathbf{L}\|_{\mathbf{R}^m} < \epsilon\,.$$

In simboli:

$$\forall\epsilon > 0\ \exists I_\epsilon(\mathbf{r_0}) \mid \forall \mathbf{r} \in \Omega \cap (\, I_\epsilon(\mathbf{r_0}) - \{\mathbf{r_0}\}\,) \Longrightarrow \|\mathbf{F}(\mathbf{r}) - \mathbf{L}\|_{\mathbf{R}^m} < \epsilon\,.$$

Se $\mathbf{L} = 0$ allora si dice che la funzione $\mathbf{F}(\mathbf{r})$ è *infinitesima* per $\mathbf{r} \to \mathbf{r_0}$.
Se accade che $\mathbf{F}(\mathbf{r})$ è definita in $\mathbf{r_0}$ e se inoltre $\lim_{\mathbf{r}\to\mathbf{r_0}} \mathbf{F}(\mathbf{r}) = \mathbf{F}(\mathbf{r_0})$, allora si dice che $\mathbf{F}(\mathbf{r})$ è *continua* in $\mathbf{r_0}$.

Si noti che la definizione di limite e quella di continuità sono formalmente analoghe a quelle note per funzioni di una variabile. la differenza importante da tener presente è che ora il simbolo $I_\epsilon(\mathbf{r}_0)$ indica una palla aperta di centro $\mathbf{r}_0$.

E' un po' diversa la definizione dei limiti infiniti o per $\mathbf{r}$ tendente ad infinito, perché non si definisce il limite "direzionale". Piuttosto, si definisce

Sia Ω un insieme illimitato. Si dice che

$$\lim_{\|\mathbf{r}\|_{\mathbf{R}^n}\to+\infty} \mathbf{F}(\mathbf{r}) = \mathbf{L}$$

se per **ogni** $\epsilon > 0$ **esiste** δ tale che se $\boxed{\|\mathbf{r}\|_{\mathbf{R}^n} > \delta}$ e $\mathbf{r} \in \Omega$ si ha

$$\boxed{\|\mathbf{F}(\mathbf{r}) - L\|_{\mathbf{R}^m} < \epsilon\,.}$$

Sia $\mathbf{r_0}$ punto di accumulazione di Ω. Si dice che

$$\lim_{\mathbf{r}\to\mathbf{r_0}} \|\mathbf{F}(\mathbf{r})\|_{\mathbf{R}^m} = +\infty$$

se per **ogni** $\epsilon > 0$ **esiste** $\delta > 0$ tale che se $\mathbf{r} \in \Omega$, $\boxed{\mathbf{r} \neq \mathbf{r_0}}$ e $\|\mathbf{r} - \mathbf{r_0}\|_{\mathbf{R}^n} < \delta$ si ha

$$\boxed{\|\mathbf{F}(\mathbf{r})\|_{\mathbf{R}^m} > \epsilon\,.}$$

In queste definizioni, per completezza, abbiamo usato gli indici $\mathbf{R}^n$ e $\mathbf{R}^m$ per indicare esplicitamente gli spazi nei quali le norme vanno lette; da ora in poi però ometteremo tali indici.

Osservazione 121 Una funzione $\mathbf{F}(\mathbf{r})$ da $\mathbf{R}^n$ in $\mathbf{R}^m$ è infinitesima (per $\mathbf{r} \to \mathbf{r_0}$ oppure per $|\mathbf{r}| \to +\infty$) se e solo se $\|\mathbf{F}(\mathbf{r})\|$ è infinitesima come funzione da $\mathbf{R}^n$ in $\mathbf{R}$. ∎

Si rileggano i teoremi sui limiti delle funzioni di una variabile, identificando quelle dimostrazioni che, grazie alle proprietà della norma, si ripetono senza alcuna modifica nel caso delle funzioni di più variabili. Naturalmente non rientrano tra queste quelle dimostrazioni che dipendono dall'ordine tra i numeri reali, come il teorema delle funzioni monotone e le sue conseguenze. Sottolineiamo che in particolare valgono:

1. Il *teorema di unicità del limite;*
2. il *teorema di limitatezza locale;*
3. i teoremi che correlano i limiti e la continuità con le operazioni. Naturalmente, tra funzioni a valori vettoriali non si fanno divisioni; e il teorema del prodotto va applicato al prodotto scalare o anche, quando $m = 3$, al prodotto vettoriale: i prodotti scalari e vettoriali di funzioni continue sono funzioni continue;
4. i teoremi relativi ai limiti di funzioni composte;
5. per funzioni a valori in $\mathbf{R}$, il *teorema di confronto per i limiti e il teorema di permanenza del segno;*
6. il *teorema di confronto per gli infiniti e gli infinitesimi.*

In particolare, vale

Lemma 122 *Sia*

$$\lim_{\mathbf{r}\to\mathbf{r_0}} \frac{\mathbf{F}(\mathbf{r})}{\|\mathbf{r}-\mathbf{r_0}\|} = 0\,.$$

Sia

$$\lim_{t\to t_0} \mathbf{r}(t) = \mathbf{r_0}$$

ed esista un intorno di t_0 su cui la funzione $\mathbf{r}(t)$ non prende valore $\mathbf{r_0}$. In tal caso vale

$$\lim_{t\to t_0} \frac{\mathbf{F}(\mathbf{r}(t))}{\|\mathbf{r}(t)-\mathbf{r_0}\|} = 0\,.$$

Notiamo un caso particolare del teorema sulla continuità delle funzioni composte, già usato nel corso di Analisi Matematica 1, trattando le equazioni differenziali:

Teorema 123 *Sia $f(\mathbf{x})$ un funzione da $\mathbf{R}^n$ in $\mathbf{R}^m$, continua su un insieme Ω. Sia $\mathbf{x}(t)$ una funzione continua della variabile reale $t \in [a,b]$, a valori in Ω. La funzione composta $f(\mathbf{x}(t))$ è continua su $[a,b]$.*

Come si è detto, una funzione che ha limite zero (per $\mathbf{r} \to \mathbf{r_0}$) si chiama ancora un *infinitesimo* (per $\mathbf{r} \to \mathbf{r_0}$); e si può istituire un confronto tra gli infinitesimi, esattamente come nel caso di funzioni di una variabile. Per esempio, il simbolo

$$\mathbf{F} = \mathrm{o}(\,\|\mathbf{r}-\mathbf{r_0}\|\,)$$

significa

$$\lim_{\mathbf{r}\to\mathbf{r_0}} \frac{\|\mathbf{F}(\mathbf{r})\|}{\|\mathbf{r}-\mathbf{r_0}\|} = 0\,.$$

Usando il linguaggio degli infinitesimi, possiamo enunciare:

Teorema 124 *Si ha*

$$\lim_{\mathbf{r}\to\mathbf{r_0}} \mathbf{F}(\mathbf{r}) = \mathbf{L}$$

se e solo se la funzione $\mathbf{F}(\mathbf{r}) - \mathbf{L}$ è infinitesima per $\mathbf{r} \to \mathbf{r_0}$.

Infine, notiamo che le disuguaglianze (3.3) permettono di provare:

Teorema 125 *Siano $F_i(\mathbf{r})$ le componenti della funzione $\mathbf{F}(\mathbf{r})$ da $\mathbf{R}^n$ in $\mathbf{R}^m$. Si ha $\lim_{\mathbf{r}\to\mathbf{r_0}} \mathbf{F}(\mathbf{r}) = \mathbf{L}$ se e solo se per ogni indice i vale $\lim_{\mathbf{r}\to\mathbf{r_0}} F_i(\mathbf{r}) = L_i$, ove L_i sono le componenti del vettore $\mathbf{L}$.*

Si enunci il risulato analogo per i limiti per $\|\mathbf{r}\| \to +\infty$.

Nonostante che abbiamo insistito sulla completa corrispondenza che intercorre tra definizioni e teoremi in una e più variabili, bisogna sottolineare una importante dissimmetria, che illustriamo con riferimento a funzioni da $\mathbf{R}^2$ in $\mathbf{R}^2$. Sia quindi

$$\mathbf{F}(x, y) = f(x, y)\mathbf{i} + g(x, y)\mathbf{j}\,.$$

Da

$$\begin{aligned} |f(x, y)| \le \|\mathbf{F}(x, y)\| \le |f(x, y)| + |g(x, y)| \\ |g(x, y)| \le \|\mathbf{F}(x, y)\| \le |f(x, y)| + |g(x, y)| \end{aligned}$$

si vede che $\mathbf{F}(x, y)$ è infinitesima se e solo se ambedue le sue componenti lo sono; è continua se e solo se ambedue le sue componenti lo sono; ossia, come asserito dal Teorema 125, limiti e continuità possono studiarsi esaminando separatamente le due componenti dei valori assunti dalla funzione, ossia le due componenti $f(x)$ ed $g(x)$ di $\mathbf{F}(x)$. Ciò non accade nello spazio di partenza; ossia le due componenti x ed y non possono trattarsi separatamente. Per vedere questo, basta considerare una sola delle componenti di $\mathbf{F}(x, y)$, per esempio la funzione a valori reali $f(x, y)$. La funzione

$$\phi(x) = \lim_{y\to y_0} f(x, y) \tag{4.1}$$

può essere definita per ogni x, e può esistere $\lim_{x\to x_0} \phi(x)$; però tale limite è in generale diverso da $\lim_{(x,y)\to(x_0,y_0)} f(x, y)$, come mostra l'esempio seguente:

Esempio 126 Sia

$$f(x, y) = \begin{cases} 1 & \text{se} \quad |x| < |y| \\ 0 & \text{se} \quad |x| \ge |y|\,. \end{cases}$$

La funzione $f(x, y)$ è priva di limite per (x, y) tendente a zero, mentre

$$\lim_{x\to 0}\left[\lim_{y\to 0} f(x, y)\right] = 0\,, \qquad \lim_{y\to 0}\left[\lim_{x\to 0} f(x, y)\right] = 1\,. \qquad \blacksquare$$

Si veda la figura 4.1.

Può anche accadere che esista

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y)$$

ma che non esista il limite che definisce $\phi(x)$ in (4.1), come mostra l'esempio seguente:

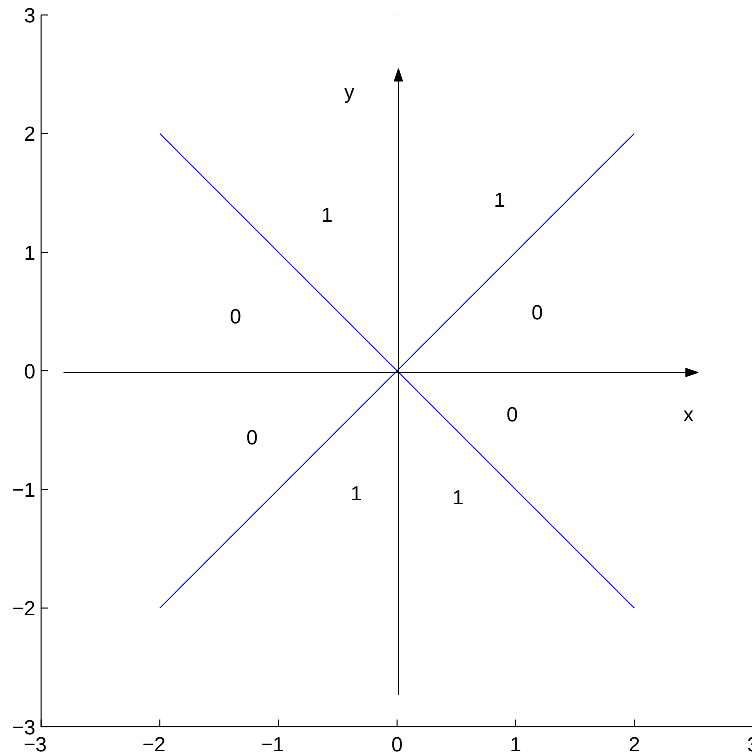
Esempio 127 Sia

$$f(x, y) = x\,[\mathrm{sgn}(y)]\,.$$

Essendo $|f(x, y)| \le |x|$, il *teorema di confronto dei limiti* mostra che

$$\lim_{(x,y)\to(0,0)} f(x, y) = 0\,;$$

Figura 4.1: La funzione dell'esempio 126



ma,

$$\lim_{y \to 0} f(x, y)$$

esiste soltanto se $x = 0$. ∎

Infine, ricordiamo che *successione* è il termine che si usa per designare una funzione definita sui numeri naturali. Una successione a valori in $\mathbf{R}^m$ si indicherà col simbolo $(\mathbf{r}_n)$ o, più frequentemente $\{\mathbf{r}_n\}$[1].

Come nel caso particolare delle successioni a valori reali, le definizioni e i teoremi sui limiti delle successioni sono casi particolari di quelli relativi alle funzioni.

4.1.1 Funzioni continue su insiemi

E' possibile provare l'analogo del Teorema di *Bolzano-Weierstrass:* ogni successione limitata a valori in $\mathbf{R}^n$ ammette sottosuccessioni convergenti.

[1]se si usa il simbolo con la parentesi graffa, si deve fare attenzione a capire quando si indica la successione oppure quando se ne vuole indicare l'insieme immagine.

Un insieme Ω si dice *compatto*, quando ogni successione in Ω ammette sottosuccessioni convergenti a punti di Ω. Si può ancora provare:

Teorema 128 *Un insieme Ω è compatto se e solo se è limitato e chiuso.*

Grazie a ciò, si prova l'analogo del *Teorema di Weierstrass* con la medesima dimostrazione vista per funzioni di una variabile.

Diremo che:

1. un punto $\mathbf{r_0}$ è *punto di minimo, o di massimo*, per una funzione $f(\mathbf{r})$ a valori reali, se vale, rispettivamente,

$$f(\mathbf{r}) \geq f(\mathbf{r_0}) \qquad \text{oppure} \qquad f(\mathbf{r}) \leq f(\mathbf{r_0})$$

per ogni $\mathbf{r} \in \Omega$;

2. diremo che la funzione $\mathbf{F}(\mathbf{r})$, da $\mathbf{R}^n$ in $\mathbf{R}^m$ è *uniformemente continua* se **per ogni** $\epsilon > 0$ esiste $\delta > 0$ tale che

$$\mathbf{r} \in \Omega\,, \quad \mathbf{r_0} \in \Omega\,, \qquad \|\mathbf{r} - \mathbf{r_0}\| < \delta \quad \text{implica} \quad \|\mathbf{F}(\mathbf{r}) - \mathbf{F}(\mathbf{r_0})\| < \epsilon\,.$$

Vale:

Teorema 129 *Sia $f(\mathbf{r})$ una funzione* a valori reali *definita e continua su un compatto $K \subseteq \mathbf{R}^n$. La funzione $f(\mathbf{r})$* ammette *in K sia punti di* minimo *che punti di* massimo.

Il teorema di esistenza degli zeri, equivalentemente quello dei valori intermedi, richiede, per funzioni di una variabile, di lavorare su intervalli. Gli insiemi aperti e connessi sono gli analoghi degli intervalli aperti, nel senso che per essi vale il teorema seguente:

Teorema 130 *Sia Ω un insieme* aperto e connesso *e sia $f(\mathbf{r})$ una funzione* a valori reali *e continua su Ω. Se $\mathbf{r_0}$ e $\mathbf{r_1}$ sono due punti di Ω, la funzione assume ogni valore c compreso tra $f(\mathbf{r_0})$ e $f(\mathbf{r_1})$.*

Dim. Sia

$$f(\mathbf{r_0}) = a\,, \qquad f(\mathbf{r_1}) = b\,.$$

Se è possibile congiungere $\mathbf{r_0}$ e $\mathbf{r_1}$ con un segmento contenuto in Ω, si considera la restrizione della funzione a tale segmento. Si trova una funzione di una sola variabile, continua e definita su un intervallo, che assume i valori a e b. Dunque assume anche il valore intermedio c.

Nel caso che i due punti si congiungano con una poligonale, ossia, con un numero finito di segmenti, si procede in modo analogo sui singoli segmenti. ∎

Esistono versioni del teorema precedente anche per funzioni a valori vettoriali. Si veda per esempio il paragrafo 7.5.2.

4.2 Le proprietà di differenziabilità

Per lo studio di queste proprietà, conviene separare lo studio delle funzioni a valori in $\mathbf{R}$ da quello delle funzioni a valori in $\mathbf{R}^m$ con $m > 1$.

4.2.1 Il differenziale delle funzioni a valori reali

Quando la funzione $f(x, y)$ dipende da due variabili, la *derivata* *parziale* rispetto ad x è la funzione sia di x che di y definita da

$$\lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} .$$

Per indicare questa funzione si usa uno dei simboli

$$\frac{\partial}{\partial x} f(x, y) , \qquad f_x(x, y) .$$

In modo analogo si definisce la derivata parziale rispetto ad y e quella rispetto alle ulteriori variabili nel caso che la funzione dipenda da più di due variabili.

Si noti esplicitamente che trattando delle equazioni differenziali si è dovuto richiedere però non la sola esistenza delle derivate parziali, ma la loro continuità. In effetti, la sola esistenza delle derivate parziali è un concetto molto debole. Infatti:

Esempio 131 L'esistenza delle derivate parziali in un punto (x_0, y_0), non implica la continuità della funzione in tale punto, come mostra l'esempio seguente:

$$f(x, y) = \begin{cases} 0 & \text{se } x \cdot y = 0 \\ 1 & \text{altrimenti,} \end{cases} \qquad (x_0, y_0) = (0, 0) .$$

■

Per contrasto, si ricordi che l'esistenza della derivata prima in un punto di una funzione di una sola variabile, implica la continuità in tale punto. Ricordiamo che la dimostrazione di ciò segue dalla prima formula degli incrementi finiti e che praticamente tutte le proprietà delle funzioni derivabili di una variabile seguono dalla prima oppure dalla seconda formula degli incrementi finiti. Dunque, se si vuol sperare di ripetere, per le funzioni di più variabili, una teoria simile a quella delle funzioni di una variabile, dovremo dare condizioni per l'esistenza di una formula analoga alla prima formula degli incrementi finiti: ossia, nel caso di funzioni di due variabili, vorremo condizioni perchè valga la formula seguente

$$f(x, y) - f(x_0, y_0) = a(x - x_0) + b(y - y_0) + \mathrm{o}(\| (x - x_0, y - y_0) \|) . \tag{4.2}$$

Il risultato che vogliamo provare è:

Teorema 132 *Sia $f(x, y)$ una funzione di due variabili. Supponiamo che essa ammetta ambedue le derivate parziali* in ogni punto (x, y) di un intorno di (x_0, y_0) *e che queste siano* continue in tale intorno. *Esistono numeri a, b per i quali vale la formula (4.2), ed è:*

$$a = f_x(x_0, y_0) , \qquad b = f_y(x_0, y_0) .$$

La dimostrazione è in appendice.

Passando al limite per $\mathrm{dist}((x, y), (x_0, y_0))$ tendente a zero in (4.2) si vede che $\lim f(x, y) = f(x_0, y_0)$. Si ha quindi in particolare:

Corollario 133 *Se una funzione ha derivate parziali* continue un intorno *di (x_0, y_0), essa è continua in (x_0, y_0).*

Argomenti del tutto analoghi valgono per funzioni di tre o più variabili:

Teorema 134 *Una funzione di n variabili le cui n derivate prime esistono e sono continue su un aperto Ω è continua su Ω e per essa vale*

$$f(\xi^1,\xi^2,\dots,\xi^n) = f(x^1,x^2,\dots,x^n) + \sum_{i=1}^n (\xi^i - x^i) f_{x^i}(x^1,x^2,\dots,x^n)$$
$$+\mathrm{o}(\|\mathbf{h}\|)\,, \qquad \mathbf{h} = \big((\xi^1-x^1),(\xi^2-x^2),\dots,(\xi^n-x^n)\big)\,. \qquad (4.3)$$

Si chiama *differenziale* della funzione f in $\mathbf{r} = \begin{bmatrix} x^1 & x^2 & \dots & x^n \end{bmatrix}'$ la trasformazione che al vettore $(y_1,\dots,y_n)$ associa il numero

$$\sum_{i=1}^n y_i f_{x^i}(x^1,x^2,\dots,x^n)\,.$$

Una funzione dotata di differenziale si chiama *differenziabile*.

Il vettore

$$\begin{bmatrix} f_{x^1}(\mathbf{r}) \\ f_{x^2}(\mathbf{r}) \\ \vdots \\ f_{x^n}(\mathbf{r}) \end{bmatrix} \qquad \mathbf{r} = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$$

si chiama il *gradiente* della funzione $f(\mathbf{r})$ e si indica col simbolo $\nabla f(\mathbf{r})$. Il simbolo ∇ si legge "grad", oppure "del" od anche "nabla".

Usando questo simbolo, la (4.3) si scrive

$$f(\mathbf{r}) - f(\mathbf{r_0}) = \nabla f(\mathbf{r_0})\cdot(\mathbf{r}-\mathbf{r_0}) + \mathrm{o}(\|\mathbf{r}-\mathbf{r_0}\|)\,.$$

Si chiama *piano tangente* al grafico di $f(\mathbf{r})$ nel punto $\mathbf{r}_0 = (x_0^1,\dots,x_0^n)$ il grafico della funzione

$$\mathbf{r} \to \mathbf{r}_0 + \nabla f(\mathbf{r}_0)\cdot(\mathbf{r}-\mathbf{r}_0)\,.$$

Dunque, l'equazione del piano tangente è

$$\mathbf{y} = \mathbf{r}_0 + \nabla f(\mathbf{r}_0)\cdot(\mathbf{r}-\mathbf{r}_0)\,.$$

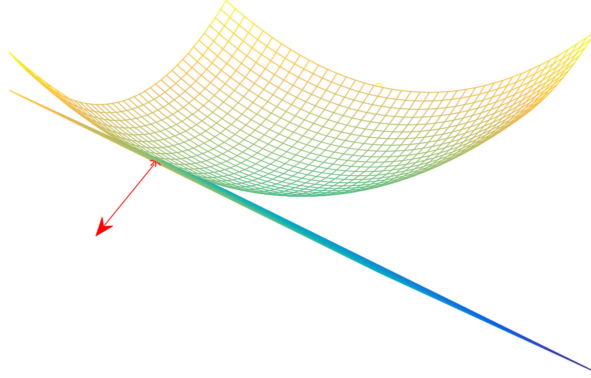
Il vettore $\nabla f(\mathbf{r}_0)$ è ortogonale al piano tangente. Per definizione, si chiama il *vettore normale* al grafico di $f(\mathbf{r})$ nel punto $(\mathbf{r}_0, f(\mathbf{r}_0))$. Si veda la figura 4.2.

Una funzione dotata di derivate parziali prime continue su un aperto Ω si chiama una funzione *di classe $C^1(\Omega)$* e si scrive $f \in C^1(\Omega)$.

Si noti che, per noi, il gradiente è un vettore colonna. D'altra parte, il differenziale, come trasformazione da $\mathbf{R}^n$, dotato della base canonica, ad $\mathbf{R}$, si rappresenta mediante una matrice $1\times n$, ossia mediante un vettore riga. Il vettore riga che rappresenta il differenziale è il trasposto del gradiente.

Diciamo infine che si chiamano *punti stazionari* di $f(\mathbf{r})$ i punti nei quali si annulla il gradiente $\nabla f(\mathbf{r})$. Tali punti si chiamano anche punti *estremali* o punti *critici*.

Figura 4.2: Piano tangente e vettore normale



4.2.2 Regole di derivazione

Dato che la derivata parziale rispetto ad x di una funzione $f(x, y)$ si calcolano fissando prima il valore di y, e lavorando con la funzione della sola x, si hanno immediatamente le regole seguente:

$$\begin{aligned}
&\frac{\partial}{\partial x} af(x, y) = a\frac{\partial}{\partial x} f(x, y) \qquad \forall a \in \mathbf{R}\,;\\
&\frac{\partial}{\partial x}\left(f(x, y) + g(x, y)\right) = \frac{\partial}{\partial x} f(x, y) + \frac{\partial}{\partial x} g(x, y)\,;\\
&\frac{\partial}{\partial x}\left(f(x, y)g(x, y)\right) = \left(\frac{\partial}{\partial x} f(x, y)\right) g(x, y) + f(x, y)\left(\frac{\partial}{\partial x} g(x, y)\right)\,;\\
&\frac{\partial}{\partial t} f(x(t), y) = \left(\frac{\partial}{\partial x} f(x(t), y)\right) x'(t)\,.
\end{aligned}$$

Però nello studio della funzione composta si incontrano casi più complessi: può essere che sia x che y vengano a dipendere dalla medesima variabile, ossia che si voglia calcolare

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x(t), y(t))\,.$$

In questo caso si ha:

Teorema 135 *Sia* $f(x, y)$ differenziabile *in ogni punto di una regione* $\Omega \subseteq \mathbf{R}^2$ *e sia* $(x(t), y(t))$ *una funzione derivabile di* $t \in (a, b)$ *a valori in* Ω. *Allora per ogni* t *si ha:*

$$\frac{d}{dt} f(x(t), y(t)) = f_x(x(t), y(t))\dot{x}(t) + f_y(x(t), y(t))\dot{y}(t)\,. \tag{4.4}$$

Dim. Si fissi un punto t_0 in cui si vuol calcolare la derivata, e si ponga $a = f_x(x(t_0), y(t_0))$, $b = f_y(x(t_0), y(t_0))$. Si ha:

$$\begin{aligned}
&f(x(t), y(t)) - f(x(t_0), y(t_0)) = a[x(t) - x(t_0)] + b[y(t) - y(t_0)] \\
&+\mathrm{o}(\| (x(t) - x(t_0), y(t) - y(t_0)) \|) \\
&= [a\dot{x}(t_0) + b\dot{y}(t_0)](t - t_0) + \mathrm{o}(t - t_0) + \mathrm{o}(\| (x(t) - x(t_0), y(t) - y(t_0)) \|) \,.
\end{aligned}$$

Notando che

$$\frac{\mathrm{o}(\| (x(t) - x(t_0), y(t) - y(t_0)) \|)}{\|(x(t) - x(t_0), y(t) - y(t_0))\|} \cdot \frac{\|(x(t) - x(t_0), y(t) - y(t_0))\|}{t - t_0}$$

si vede che

$$\mathrm{o}(\| (x(t) - x(t_0), y(t) - y(t_0)) \|) = \mathrm{o}(t - t_0) \,.$$

Infatti

$$\frac{\mathrm{o}(\| (x(t) - x(t_0), y(t) - y(t_0)) \|)}{\|(x(t) - x(t_0), y(t) - y(t_0))\|}$$

tende a zero per il Lemma 122 mentre

$$\frac{\|(x(t) - x(t_0), y(t) - y(t_0))\|}{t - t_0}$$

rimane limitata, come si vede dalla seconda formula degli incrementi finiti applicata sia ad $x(t)$ che ad $y(t)$:

$$\frac{\|(x(t) - x(t_0), y(t) - y(t_0))\|}{t - t_0} = \sqrt{(\dot{x}(c))^2 + (\dot{y}(d))^2} \,.$$

Sia ora

$$g(t) = f(x(t), y(t)) \,.$$

L'uguaglianza

$$\begin{aligned}
&g(t) - g(t_0) = [a\dot{x}(t_0) + b\dot{y}(t_0)](t - t_0) + \mathrm{o}(t - t_0) \\
&+\mathrm{o}(\| (x(t) - x(t_0), y(t) - y(t_0)) \|) \\
&= [a\dot{x}(t_0) + b\dot{y}(t_0)](t - t_0) + \mathrm{o}(t - t_0)
\end{aligned}$$

mostra che $g(t)$ è derivabile per $t = t_0$, con

$$\dot{g}(t) = \frac{\mathrm{d}}{\mathrm{d}t} f(x(t), y(t)) = a\dot{x}(t_0) + b\dot{y}(t_0) \,. \quad \blacksquare$$

Osservazione 136 Si noti che usando il Lemma 122, si è implicitamente assunto che in un opportuno intorno di t_0 la funzione $(x(t), y(t))$ non prenda valore $(x(t_0), y(t_0))$. Si provi per esercizio che questa condizione può rimuoversi. ■

Naturalmente, il Teorema 135 si estende al caso di funzioni di n variabili e, ricordando il Teorema 134, possiamo enunciare:

Corollario 137 *Sia $f(\mathbf{r}) \in C^1(\Omega)$ e sia $\mathbf{r}(t)$ una funzione di $t \in (a,b)$ a valori in Ω, derivabile. Allora, la funzione composta $f(\mathbf{r}(t))$ è derivabile su (a,b) ed inoltre:*

$$\frac{d}{dt} f(\mathbf{r}(t)) = \nabla f(\mathbf{r}(t)) \cdot \dot{\mathbf{r}}(t)$$

(il simbolo $\cdot$ indica il prodotto scalare).

Sia ora $x = x(u,v)$, $y = y(u,v)$. Applicando il teorema precedente alla variabile u, con v fissato, e quindi alla variabile v, con u fissato si trova:

Teorema 138 *Sia $f(x,y) \in C^1(\Omega)$. Siano $x = x(u,v)$, $y = y(u,v)$ due funzioni definite su $\tilde{\Omega}$ e di classe $C^1(\tilde{\Omega})$, a valori in Ω. Valgono le uguaglianze*

$$\begin{aligned} &\frac{\partial}{\partial u} f(x(u,v), y(u,v)) = f_x(x(u,v), y(u,v)) x_u(u,v) + f_y(x(u,v), y(u,v)) y_u(u,v) \\ &= \nabla f(x(u,v), y(u,v)) \cdot \begin{bmatrix} x_u(u,v) \\ y_u(u,v) \end{bmatrix}, \\ &\frac{\partial}{\partial v} f(x(u,v), y(u,v)) = f_x(x(u,v), y(u,v)) x_v(u,v) + f_y(x(u,v), y(u,v)) y_v(u,v) \\ &= \nabla f(x(u,v), y(u,v)) \cdot \begin{bmatrix} x_v(u,v) \\ y_v(u,v) \end{bmatrix} \end{aligned}$$

(il punto indica il prodotto scalare di vettori colonna).

Questa formula si chiama anche *formula della derivazione a catena*.

Consideriamo ora un caso particolare: sia $x = tv^1$, $y = tv^2$. In questo caso, i punti $(x_0 + tv^1, y_0 + tv^2)$ sono punti di una retta nella direzione del vettore $\mathbf{v} = (v^1, v^2)$, uscente da (x_0, y_0). La derivata

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x_0 + tv^1, y_0 + tv^2)$$

calcolata per $t = 0$ si chiama la *derivata secondo il vettore* $\mathbf{v}$ della funzione e si indica col simbolo

$$\frac{\partial f}{\partial \mathbf{v}}(x,y)\,.$$

Dal Teorema 135 segue che, se $f(x,y)$ è di classe C^1,

$$\frac{\partial f}{\partial \mathbf{v}}(x,y) = f_x(x,y)v^1 + f_y(x,y)v^2 = \nabla f(x,y) \cdot \mathbf{v}\,.$$

Se $\mathbf{v}$ è un versore, allora si parla di *derivata direzionale* nella direzione $\mathbf{v}$.

E' facile estendere gli argomenti precedenti a funzioni di tre o più variabili.

4.2.3 La direzione del gradiente e la direzione di massima velocità crescita

Sia $f(\mathbf{r})$ una funzione differenziabile a valori reali e sia $\mathbf{r_0}$ un punto del suo dominio, che supponiamo aperto. Fissiamo un segmento uscente da $\mathbf{r_0}$, dato da

$$\mathbf{r} = \mathbf{r_0} + t\mathbf{v}\,, \qquad t \in (-\epsilon, \epsilon)\,. \tag{4.5}$$

Il numero ϵ si sceglie piccolo, in modo che il segmento sia contenuto nel dominio della funzione, e $\mathbf{v}$ è un versore.

La derivata in $t = 0$ della funzione $f(\mathbf{r_0} + t\mathbf{v})$ rappresenta la velocità di variazione dei valori della restrizione di $f(\mathbf{r})$ al segmento (4.5). E'

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{r_0} + t\mathbf{v})_{|t=0} = \nabla f(\mathbf{r_0}) \cdot \mathbf{v}\,.$$

Questo numero rappresenta la componente di $\nabla f(\mathbf{r_0})$ lungo $\mathbf{v}$ e quindi il suo valore assoluto è sempre minore di $\|\nabla f(\mathbf{r_0})\|$; è uguale quando accade che $\mathbf{v}$ è il versore

$$\mathbf{v} = \frac{\nabla f(\mathbf{r_0})}{\|\nabla f(\mathbf{r_0})\|}\,.$$

In questo caso si ha anche

$$\nabla f(\mathbf{r_0}) \cdot \mathbf{v} = \|\nabla f(\mathbf{r_0})\|\,;$$

ossia, la funzione $f(\mathbf{r})$ ha massima velocità di crescita nella direzione e verso di $\nabla f(\mathbf{r_0})$; nel verso opposto essa ha massima velocità di decrescita.

4.2.4 Le funzioni definite tramite integrali

Sia $f(x, y)$ una funzione continua su un rettangolo chiuso $[a, b] \times [c, d]$. Per ogni fissata coppia di numeri $x \in [a, b]$ e $y \in [c, d]$ si può definire il numero

$$\phi(x, y) = \int_c^y f(x, s)\,\mathrm{d}s\,.$$

Si trova così una seconda funzione definita su $[a, b] \times [c, d]$. Vale:

Teorema 139 *La funzione $f(x, y)$ sia continua sul rettangolo chiuso R. Allora, la funzione $\phi(x, y)$ è continua in R.*

Dim. Si scriva

$$|\phi(x, y) - \phi(x', y')| \le |\phi(x, y) - \phi(x', y)| + |\phi(x', y) - \phi(x', y')|\,.$$

Vogliamo provare che **per ogni** $\eta > 0$ **esiste** $\sigma > 0$ tale che se

$$d((x, y), (x', y')) < \sigma$$

allora vale

$$|\phi(x, y) - \phi(x', y)| < \eta/2\,, \tag{4.6}$$
$$|\phi(x', y) - \phi(x', y')| < \eta/2\,. \tag{4.7}$$

Studiamo (4.6):

$$\begin{aligned}
|\phi(x, y) - \phi(x', y)| &= \left| \int_c^y f(x, s)\,\mathrm{d}s - \int_c^y f(x'.s)\,\mathrm{d}s \right| \\
&= \left| \int_c^y [f(x, s) - f(x', s)]\ \mathrm{d}s \right| \le \int_c^y |f(x, s) - f(x', s)|\ \mathrm{d}s \\
&\le \int_c^d |f(x, s) - f(x', s)|\ \mathrm{d}s\,.
\end{aligned}$$

La funzione $f(x, y)$ è continua sul rettangolo limitato e chiuso R e quindi è uniformemente continua: per ogni $\delta > 0$ esiste $\sigma > 0$ tale che

$$\|(x, y) - (x', y')\| < \sigma \implies |f(x, y) - f(x', y')| < \delta\,.$$

La condizione

$$\|(x, y) - (x', y')\| < \sigma$$

vale in particolare se $y = y' \in [c, d]$ e se $|x - x'| < \sigma$. Dunque,

$$|x - x'| < \sigma \implies |f(x, s) - f(x', s)| < \delta \implies \int_c^d |f(x, s) - f(x', s)|\, \mathrm{d}s < (d - c)\delta\,.$$

L'assero segue scegliendo

$$\delta = \frac{\eta}{2(d - c)}\,.$$

Studiamo il secondo addendo (4.7). Per fissare le idee sia $y < y'$:

$$|\phi(x', y) - \phi(x', y')| = \left| \int_c^y f(x', s)\, \mathrm{d}s - \int_c^{y'} f(x', s)\, \mathrm{d}s \right| \le \int_y^{y'} |f(x', s)|\, \mathrm{d}s\,.$$

La funzione $f(x, y)$ è continua sul rettangolo limitato e chiuso R. Dunque, per il *Teorema di Weierstrass,* è limitata. Sia $|f(x, y)| < M$. Per la monotonia dell'integrale si ha

$$|\phi(x', y) - \phi(x', y')| \le M(y' - y)\,.$$

Per avere questa differenza minore di $\eta/2$ basta imporre la condizione $|y' - y| < \eta/2M$. ∎

Di conseguenza, dal *teorema sulla continuità delle funzioni composte,* viene ad essere continua anche la funzione

$$\int_c^{g(x,y)} f(x, s)\, \mathrm{d}s\,,$$

per ogni funzione continua $g(x, y)$. Scegliendo in particolare $g(x, y)$ costantemente uguale a b si trova:

Teorema 140 *Sia $f(x, y)$ continua su $[a, b] \times [c, d]$ e sia*

$$\phi(x) = \int_a^b f(x, y)\, dy\,.$$

La funzione $\phi(x)$ è continua su $[a, b]$.

Studiamo ora la derivabilità:

Teorema 141 *Il rettangolo chiuso R sia interno ad una regione Ω su cui la funzione $f(x, y)$ ammette derivate parziali continue. Si ha:*

$$\frac{\partial}{\partial x}\phi(x, y) = \frac{\partial}{\partial x}\int_c^y f(x, s)\, ds = \int_c^y f_x(x, s)\, ds$$
$$\frac{\partial}{\partial y}\phi(x, y) = \frac{\partial}{\partial y}\int_c^y f(x, s)\, ds = f(x, y)\,.$$

Dim. La seconda uguaglianza è ovvia perchè il calcolo dell'integrale e quello della derivata parziale rispetto ad y si fanno per ogni x fissata; e quindi di fatto si lavora con funzioni della sola variabile y.

Proviamo la prima uguaglianza. Scriviamo

$$\frac{\phi(x+h,y)-\phi(x,y)}{h}-\int_c^y f_x(x,s)\,\mathrm{d}s=\int_c^y\left[\frac{f(x+h,s)-f(x,s)}{h}-f_x(x,s)\right]\mathrm{d}s\,.$$

Vogliamo provare che questa differenza tende a zero per $h\to 0$. Fissiamo $s\in[c,d]$ e scriviamo la formula della media per la funzione della sola variabiled x:

$$\frac{f(x+h,s)-f(x,s)}{h}-f_x(x,s)=f_x(\tilde{x},s)-f_x(x,s)$$

ove $\tilde{x}$ dipende sia da h che da s. E' però un punto dell'intervallo $(x,x+h)$.

La funzione $f_x(x,y)$ è uniformemente continua su R e quindi, dato $\epsilon>0$, esiste $\delta_\epsilon>0$ tale che

$$\text{se } |h|<\delta_\epsilon \quad \text{vale} \quad |f_x(\tilde{x},s)-f_x(x,s)|<\epsilon\,;$$

e quindi,

$$\left|\frac{\phi(x+h,y)-\phi(x,y)}{h}-\int_c^y f_x(x,s)\,\mathrm{d}s\right|\leq\int_c^y \epsilon\,\mathrm{d}s\leq(d-c)\epsilon\,.$$

Ciò prova che il limite per h tendente a zero è nullo. ∎

Torniamo ora a considerare la funzione

$$\phi(x)=\int_c^d f(x,y)\,\mathrm{d}y\,.$$

Essendo questa funzione continua, essa può venir integrata rispetto alla variabile x:

$$\int_a^b\left[\int_c^d f(x,y)\,\mathrm{d}y\right]\,\mathrm{d}x$$

Si chiama questo *integrale* *iterato* della funzione $f(x,y)$. Naturalmente, si può anche introdurre un secondo integrale iterato,

$$\int_c^d\left[\int_a^b f(x,y)\,\mathrm{d}x\right]\,\mathrm{d}y$$

e si pone il problema di sapere se i loro valori coincidano o meno. La risposta affermativa si vedrà nel cap. 7.

4.3 Le derivate di ordine superiore

Ricordiamo che le derivate parziali sono a loro volta funzioni di più variabili, e quindi può accadere che esse siano ulteriormente derivabili. Si possono quindi definire le funzioni

$$f_{xx}(x,y)=\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x}f(x,y)\right]\,,\quad f_{xy}(x,y)=\frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}f(x,y)\right]\quad f_{yy}(x,y)=\frac{\partial}{\partial y}\left[\frac{\partial}{\partial y}f(x,y)\right]$$

e così via.

Quando una funzione ammette tutte le derivate parziali continue fino all'ordine k incluso su un aperto Ω, si dice che essa è di classe $C^k(\Omega)$.

La notazione

$$\frac{\partial^n}{\partial x^n} f(x,y)$$

indica la *derivata n–ma* di $f(x,y)$ rispetto alla variabile x, ossia la derivata ottenuta tenendo fisso il valore di y e derivando n volte rispetto ad x. Significato analogo ha il simbolo

$$\frac{\partial^n}{\partial y^n} f(x,y)\,.$$

Le cose sono più complicate se si vogliono le *derivate miste*, ossia le derivate ottenute derivando alcune volte rispetto ad x e anche rispetto ad y perché in generale dovremo tener conto dell'ordine con cui si eseguono le derivate. Per esempio, in generale

$$D_x\left(D_y f(x,y)\right) = \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y} f(x,y)\right] \neq \frac{\partial}{\partial y}\left[\frac{\partial}{\partial x} f(x,y)\right] = D_y\left(D_x f(x,y)\right)\,.$$

Fortunatamente, ciò non avviene nei casi più interessanti per le applicazioni. Vale infatti:

Teorema 142 (di Schwarz) *Sia $f(x,y)$ una funzione delle due variabili (x,y), di classe C^1 un intorno V di (x_0,y_0). Supponiamo che nei punti di V esistano le derivate miste $D_x\left(D_y f(x,y)\right)$ ed $D_y\left(D_x f(x,y)\right)$, e supponiamo che queste siano continue. Allora, per ogni (x,y) vale*

$$D_x\left(D_y f(x,y)\right) = D_y\left(D_x f(x,y)\right)\,.$$

Esaminando la dimostrazione, in appendice, si vede facilmente che il teorema si estende al caso di funzioni di più di due variabili.

Grazie a questo teorema, le due derivate miste rispetto ad x ed y di $f(x,y)$ si indicano semplicemente con i simboli

$$f_{x,y}(x,y)\,, \qquad \frac{\partial^2}{\partial x \partial y} f(x,y)\,,$$

senza preoccuparsi dell'ordine di derivazione.

Il Teorema di Schwarz si estende a funzioni di n variabili, ed a qualsiasi ordine di derivazione, come segue:

Corollario 143 *Sia $f(\mathbf{r}) \in C^n(\Omega)$. Si derivi la funzione k_1 volte rispetto alla prima componente di $\mathbf{r}$; k_2 rispetto alla seconda ecc., ma con*

$$k_1 + k_2 + \ldots + k_n \leq n\,.$$

La derivata che si ottiene non dipende dall'ordine col quale si eseguono le derivate.

Indicare derivate successive di funzioni di più variabili è alquanto noioso. Un simbolo comodo si ottiene in questo modo. Sia $f(\mathbf{r})$ una funzione di $\mathbf{r} \in \mathbf{R}^n$. Chiamiamo *multiindice* un vettore $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ le cui componenti sono numeri interi nulli o positivi. Indichiamo con $|\alpha|$ la *lunghezza del multiindice*:

$$|\alpha| = \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n\,.$$

Col simbolo $\mathbf{r}^\alpha$ intendiamo:

$$\mathbf{r}^\alpha = (r_1^{\alpha_1}, r_2^{\alpha_2}, \dots, r_n^{\alpha_n})\,.$$

Col simbolo

$$D^\alpha f(\mathbf{r}) = \frac{\partial^{|\alpha|}}{\partial \mathbf{r}^\alpha} f(\mathbf{r})$$

si intende la derivata di $f(\mathbf{r})$ che si ottiene derivando la funzione $f(\mathbf{r})$ nell'ordine, prima α_1 volte rispetto alla prima variabile, poi α_2 volte rispetto alla seconda variabile, ecc. (intendendo che se $\alpha_i = 0$ la corrispondente derivata non si esegue).

Il Teorema di Schwarz assicura che se ciascuna derivata di ordine $|\alpha|$ di $f(\mathbf{r})$ è continua allora le derivate miste di ordine $|\alpha|$ non dipendono dall'ordine con cui vengono calcolate. Ciò giustifica la notazione $f \in C^k(\Omega)$, che si usa quando la funzione $f(\mathbf{r})$ ammette tutte le derivate di ordine k su Ω, e queste sono continue.

4.3.1 La formula di Taylor per le funzioni a valori reali

Così come nel caso delle funzioni di una variabile, gli argomenti che hanno condotto alla prima formula degli incrementi finiti possono iterarsi se la funzione ammette continue le successive N derivate. Senza entrare nei dettagli della dimostrazione (simile a quella del Teorema 132), limitiamoci a dare la formula che si ottiene nel caso delle derivate seconde. Quando la funzione dipende da due variabili, si trova

$$\begin{aligned}
&f(x,y) = f(x_0,y_0)\\
&+ [f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)]\\
&+\frac{1}{2}\left\{f_{xx}(x_0,y_0)(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0)\right.\\
&\left.+f_{yx}(x_0,y_0)(x-x_0)(y-y_0) + f_{yy}(x_0,y_0)(y-y_0)^2\right\} + R(x,y)\,.
\end{aligned}$$

e

$$\lim_{d((x,y),(x_0,y_0))\to 0} \frac{R(x,y,)}{[d((x,y),(x_0,y_0))]^2} = 0\,, \qquad \text{ossia} \quad R(x,y) = \mathrm{o}\left(\|\,(x-x_0, y-y_0)\,\|^2\right)\,. \tag{4.8}$$

Grazie al Teorema di Schwarz, la formula precedente si può scrivere

$$\begin{aligned}
&f(x,y) = f(x_0,y_0)\\
&+ [f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)]\\
&+\frac{1}{2}\left\{f_{xx}(x_0,y_0)(x-x_0)^2 + 2f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + f_{yy}(x_0,y_0)(y-y_0)^2\right\}\\
&+R(x,y)\,.
\end{aligned}$$

Questa formula si chiama ancora *formula di Taylor* arrestata al secondo ordine perché il *resto* $R(x,y)$ verifica la condizione (4.8).

Naturalmente, aggiungendo i termini con le derivate rispetto ad una terza variabile z si trova la formula di Taylor in tre variabili e, in generale, in n variabili. Si immagina facilmente che le formule divengano via via più complesse, in particolare se la funzione dipende da più di due variabili e si vuole scrivere la formula di Taylor di ordine maggiore di 2. E' però

possibile scrivere queste formule in modo compatto come segue. Riguardiamo il caso della formula di Taylor di $f(x,y)$ arrestata all'ordine 2. Si noti che il termine

$$f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)$$

si può pensare ottenuto come segue:

$$\frac{1}{1!}\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]f(x,y)=\frac{1}{1!}\left[f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)\right]\,.$$

Consideriamo ora il termine di ordine 2:

$$\frac{1}{2}\left\{f_{xx}(x_0,y_0)(x-x_0)^2+2f_{xy}(x_0,y_0)(x-x_0)(y-y_0)+f_{yy}(x_0,y_0)(y-y_0)^2\right\}$$

si può ottenere con questa regola mnemonica: si "calcola"

$$\frac{1}{2!}\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]^2$$

scrivendo

$$\frac{1}{2!}\left[(x-x_0^2)\frac{\partial^2}{\partial x^2}+2(x-x_0)(y-y_0)\frac{\partial^2}{\partial x\partial y}+(y-y_0^2)\frac{\partial^2}{\partial y^2}\right]\,.$$

Si "applica" quindi questo alla funzione $f(x,y)$ scrivendo

$$\begin{aligned}&\frac{1}{2!}\left[(x-x_0^2)\frac{\partial^2}{\partial x^2}+2(x-x_0)(y-y_0)\frac{\partial^2}{\partial x\partial y}+(y-y_0^2)\frac{\partial^2}{\partial y^2}\right]f(x,y)\\&=\frac{1}{2!}\left[(x-x_0)^2f_{xx}(x_0,y_0)+2(x-x_0)(y-y_0)f_{xy}(x_0,y_0)+(y-y_0)^2f_{xx}(x_0,y_0)\right]\,.\end{aligned}$$

Questa "regola mnemonica" si estende al caso dell'ordine maggiore di 2 e fornisce l'espressione corretta della formula di Taylor arrestata all'ordine k:

$$f(x,y)=\sum_{j=0}^{k}\frac{1}{j!}\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]^j f(x,y)+\mathrm{o}\left(\sqrt{(x-x_0)^2+(y-y_0)^2}\right)^k\,.$$

In questa formula si intende che

$$\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]^0 f(x,y)=f(x_0,y_0)\,.$$

L'estensione di questa formula al caso di funzioni dipendenti da più di due variabili è ovvia.

4.4 Gli estremi

La formula di Taylor, arrestata al primo ordine, ha già dato un'informazione importante. Infatti, ha condotto a provare che ogni funzione le cui derivate parziali sono ovunque continue è essa stessa continua. Daltra parte, nel caso delle funzioni di una sola variabile, la formula di Taylor si usa per dedurre regole che possono condurre ad identificare i punti di massimo o di minimo. Si può immaginare che anche per funzioni di più variabili si possa fare uno studio analogo. Per renderci conto di ciò, repetiamo la definizione degli estremi:

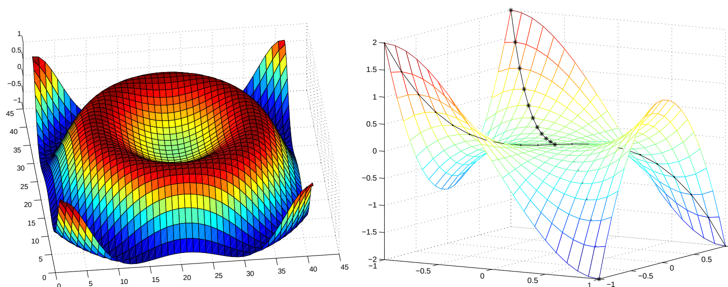


Figura 4.3: Un punto di minimo e un punto di sella

Definitione 2 Sia $f(\mathbf{r})$ una funzione da $\mathbf{R}^n \to \mathbf{R}$. Un punto $\mathbf{r_0}$ interno al dominio di $f(\mathbf{r})$ si dice *punto di* *minimo* (relativo) per la funzione $f(\mathbf{r})$ se esiste $\nu > 0$ tale che

$$d(\mathbf{r}, \mathbf{r_0}) < \nu \quad \Longrightarrow \quad f(\mathbf{r_0}) \le f(\mathbf{r})\,.$$

In modo analogo si definiscono i *punti di* *massimo*. ∎

La figura 4.3 (a sinistra) mostra un punto di minimo.

Quando l'intorno di $\mathbf{r_0}$ coincide con Ω, si parla di *estremi assoluti* o *estremi globali*. E naturalmente a questo caso ci si può ricondurre, sostituendo Ω con tale intorno.

Studiamo ora il caso degli estremi relativi di funzioni di due variabili, che sono interni al dominio.

Notiamo che se (x_0, y_0) è punto di minimo, allora la funzione $g(x)$ della sola variabile x,

$$g(x) = f(x, y_0)$$

ha minimo in x_0 ed x_0 è punto interno al dominio. Dunque, se è derivabile, essa ha derivata nulla in x_0, per il Teorema di Fermat. Per definizione, $g'(x_0) = f_x(x_0, y_0)$.

Se (x_0, y_0) è punto di minimo, esso è anche punto di minimo per la funzione

$$h(y) = f(x_0, y)$$

e quindi $h'(y_0) = 0$; e naturalmente il discorso si ripete se (x_0, y_0) è punto di massimo per $f(x, y)$. Dunque:

Teorema 144 *Se la funzione $f(x, y)$ ammette derivate parziali prime nel punto di minimo (o di massimo) (x_0, y_0) interno al dominio, esse sono ambedue nulle:*

$$f_x(x_0, y_0) = 0\,, \qquad f_y(x_0, y_0) = 0\,. \quad ∎$$

Ossia: gli estremi di una funzione vanno ricercati tra i suoi punti stazionari.

Naturalmente, un punto stazionario, ossia un punto nel quale si annulla il gradiente della funzione, non è necessariamente un estremo. Una coindizione sufficiente per gli estremi si vedrà più avanti. Però, in generale la funzione può avere un comportamente quanto mai vario nei punti nei quali si annulla il gradiente. In particolare definiamo:

Definitione 3 *Si dice che $\mathbf{r_0}$ è punto di* sella *se è un punto stazionario e se, inoltre, la funzione $f(\mathbf{r}) - f(\mathbf{r}_0)$ cambia segno in ogni intorno $\mathbf{r_0}$.* ∎

La figura 4.3, a destra, illustra un punto di sella.

Dunque il piano tangente in un punto di minimo o di massimo o di sella (interno al dominio) è orizzontale e, se la funzione è di classe C^2, può scriversi:

$$f(x,y) = f(x_0,y_0) + \frac{1}{2}\left\{f_{xx}(x_0,y_0)(x-x_0)^2 \right.$$
$$\left. +2f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + f_{yy}(x_0,y_0)(y-y_0)^2\right\} + R(x,y)$$

con $R(x,y)/[(x-x_0)^2 + (y-y_0)^2]$ infinitesima. Per capire se è possibile ripetere, per le funzioni di più variabili, uno studio analogo a quello delle funzioni di una variabile, consideriamo il caso particolare in cui $x_0 = 0$, $y_0 = 0$, $f_{xx}(0,0) = 1$, $f_{xy}(0,0) = 0$, $f_{yy}(0,0) = 1$. In questo caso,

$$\begin{aligned} f(x,y) &= f(0,0) + \frac{1}{2}\left\{f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy \right. \\ &\left. +f_{yy}(0,0)y^2\right\} + R(x,y) \\ &= f(0,0) + (x^2+y^2)\cdot\left(\frac{1}{2} + \frac{R(x,y)}{x^2+y^2}\,.\right) \end{aligned}$$

Il *teorema della permanenza del segno* mostra che per $x^2 + y^2$ abbastanza piccolo,

$$1 + \frac{1}{2}\frac{R(x,y)}{x^2+y^2} > 0$$

e quindi $f(x,y)$ ha, in $(0,0)$, un punto di minimo relativo, così come la funzione di confronto $g(x,y) = x^2 + y^2$.

In generale se $f(x,y)$ è di classe C^2 e le sue derivate prime sono ambedue nulle in un punto (x_0,y_0) dalla *formula di Taylor* si ha

$$f(x,y) = f(x_0,y_0) + g(x-x_0, y-y_0) + \mathrm{o}(\|\,(x-x_0, y-y_0)\,\|^2)$$

e la funzione di confronto $g(x,y)$ ha forma

$$\begin{aligned} g(x,y) &= \frac{1}{2}\{f_{xx}(x_0,y_0)(x-x_0)^2 + 2f_{xy}(x_0,y_0)(x-x_0)(y-y_0) \\ &+f_{yy}(x_0,y_0)(y-y_0)^2\}\,. \end{aligned}$$

Le proprietà che il punto (x_0,y_0) ha per le funzioni

$$g(x,y) = a(x-x_0)^2 + 2b(x-x_0)(y-y_0) + c(y-y_0)^2$$

sono note dai corsi di geometria, caratterizzate mediante la matrice

$$H = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Con una dimostrazione del tutto analoga a quella vista sopra in un caso particolare, si prova: Se questa matrice è definita positiva, il punto è di minimo; e tale proprietà è

ereditada dalla funzione $f(x,y)$; se la matrice è indefinita il punto è di sella, e tale proprietà è ereditata dalla funzione $f(x,y)$; se la matrice è definita negativa, il punto è di massimo; e tale proprietà è ereditada dalla funzione $f(x,y)$. Se invece la matrice è semidefinita positiva oppure semidefinita negativa, NIENTE PUÒ DIRSI SULLA FUNZIONE $f(x,y)$.

Ricordando i criteri introdotti nei corsi di geometria, per lo studio della definitezza delle matrici, si ha:

- se $a > 0$ e $\det H > 0$ la matrice H è definita positiva, ed il punto è di minimo.
- se $a < 0$ e $\det H > 0$ si ha un punto di minimo.
- se $ac = 0$ ma $\det H \neq 0$ la matrice è indefinita ed il punto è di sella.
- se $\det H = 0$ la funzione $g(x,y)$ ha per grafico un cilindro parabolico e niente può dirsi della funzione $f(x,y)$.

Questo risultato può estendersi a funzioni di n variabili. Introduciamo per questo la *matrice* *hessiana* della funzione $f(\mathbf{r})$, $\mathbf{r} = (x^1, \ldots, x^n)$:

$$H(\mathbf{r}) = \begin{bmatrix} f_{x^1,x^1}(\mathbf{r}) & f_{x^1,x^2}(\mathbf{r}) & \ldots & f_{x^1,x^n}(\mathbf{r}) \\ f_{x^2,x^1}(\mathbf{r}) & f_{x^2,x^2}(\mathbf{r}) & \ldots & f_{x^2,x^n}(\mathbf{r}) \\ \vdots & & & \\ f_{x^n,x^1}(\mathbf{r}) & f_{x^n,x^2}(\mathbf{r}) & \ldots & f_{x^n,x^n}(\mathbf{r}) \end{bmatrix}.$$

Vale:

Teorema 145 *Siano nulle le derivate parziali prime di $f(\mathbf{r})$ in $\mathbf{r_0}$. In tal caso:*

- *se la matrice hessiana è definita* positiva, *il punto è di* minimo *per $f(\mathbf{r})$;*
- *se la matrice hessiana è* indefinita *il punto è di* sella *per $f(\mathbf{r})$;*
- *se la matrice hessiana è definita* negativa, *il punto è di* massimo *per $f(x,y)$.*

Invece, NIENTE PUÒ DIRSI *se la matrice hessiana è soltanto* semidefinita *positiva oppure negativa.*

Osservazione 146 E' possibile dare ulteriori condizioni necessarie che devono essere soddisfatte nei punti di minimo oppure di massimo. Ricordiamo che in un punto di minimo $\mathbf{r_0} = (x_0, y_0)$ le derivate parziali prime devono essere tutte nulle. Questa condizione necessaria si ottiene facilmente considerando la restrizione della funzione ad una delle rette coordinate passanti per il punto; ossia, considerando la funzione $f(x, y_0)$ si trova che deve essere $f_x(x_0, y_0) = 0$.

Se esiste la derivata seconda, $f_{xx}(x_0, y_0)$ non può essere negativa, altrimenti la funzione $f(x, y_0)$ avrebbe in x_0 un punto di massimo, invece che di minimo. Dunque, in un punto di minimo (x_0, y_0) si deve avere

$$f_{xx}(x_0,y_0) \geq 0\,, \quad f_{yy}(x_0,y_0) \geq 0 \qquad \text{e quindi anche } f_{xx}(x_0,y_0) + f_{yy}(x_0,y_0) \geq 0.$$

In un punto di massimo deve aversi invece

$$f_{xx}(x_0,y_0) \leq 0\,, \quad f_{yy}(x_0,y_0) \leq 0 \qquad \text{e quindi anche } f_{xx}(x_0,y_0) + f_{yy}(x_0,y_0) \leq 0.$$

Questa condizione si generalizza immediatamente al caso di funzioni dipendenti da più variabili.

Un'ulteriore condizione ancora solamente necessaria ma più precisa è la seguente: In un punto di minimo le derivate parziali prime sono nulle e inoltre la matrice hessiana è definita positiva oppure semidefinita positiva.

Infatti, se la matrice hessiana fosse definita negativa si avrebbe un punto di massimo, se fosse indefinita si avrebbe un punto di sella.

Analogamente, In un punto di massimo le derivate parziali prime sono nulle e inoltre la matrice hessiana è definita negativa oppure semidefinita negativa. ∎

4.5 Il differenziale delle funzioni a valori in $\mathbf{R}^m$

Sia $\mathbf{F}(\mathbf{r})$ una funzione da $\mathbf{R}^n$ ad $\mathbf{R}^m$ e sia $\mathbf{J}(\mathbf{r})$ una trasformazione lineare da $\mathbf{R}^n$ ad $\mathbf{R}^m$. Si dice che la trasformazione lineare $\mathbf{J}(\mathbf{r})$ è il *differenziale* di $\mathbf{F}(\mathbf{r})$ in $\mathbf{r_0}$ se

$$\|\mathbf{F}(\mathbf{r}) - \mathbf{F}(\mathbf{r_0}) - \mathbf{J}(\mathbf{r} - \mathbf{r_0})\| = \mathrm{o}\left(\|\mathbf{r} - \mathbf{r_0}\|\right) .$$

Ovviamente, la i–ma componente $F_i(\mathbf{r})$ è una funzione a valori reali e dalla disuguaglianza

$$|F^i| \leq \|\mathbf{F}\| \qquad \text{se} \qquad \mathbf{F} = (F^1, F^2, \dots, F^m)$$

segue

$$|F^i(\mathbf{r}) - F^i(\mathbf{r_0}) - J^i(\mathbf{r} - \mathbf{r_0})\| = \mathrm{o}(\|\mathbf{r} - \mathbf{r_0}\|) .$$

ove $J^i(\mathbf{r})$ è la i–ma componente di $\mathbf{J}(\mathbf{r})$. Segue che una funzione da $\mathbf{R}^n$ in $\mathbf{R}^m$ che è differenziabile ha differenziabile, e quindi continua, ciascuna delle sue componenti:

Teorema 147 *Se la funzione $\mathbf{F}(\mathbf{r})$ è differenziabile in $\mathbf{r_0}$, essa è ivi continua.*

Viceversa, dalle disuguaglianze (3.3), esiste un M per cui

$$\|\mathbf{F}(\mathbf{r}) - \mathbf{F}(\mathbf{r_0}) - \mathbf{J}(\mathbf{r} - \mathbf{r_0})\| \leq M \sum_{i=1}^{m} \|F^i(\mathbf{r}) - F^i(\mathbf{r_0}) - J^i(\mathbf{r} - \mathbf{r_0})\| .$$

Dunque:

Teorema 148 *La funzione $\mathbf{F}(\mathbf{r})$ è differenziabile in $\mathbf{r_0}$ se e solo se ciascuna delle sue componenti lo è.*

Ricordiamo che le trasformazioni lineari da $\mathbf{R}^n$ ad $\mathbf{R}^m$ (che rappresentiamo rispetto alle basi canoniche) si rappresentano mediante matrici. Sia J la matrice della trasformazione lineare $\mathbf{J}(\mathbf{r})$, differenziale in $\mathbf{r_0}$ della funzione $\mathbf{F}(\mathbf{r})$. La sua i–ma riga rappresenta il differenziale della componente $F^i(\mathbf{r})$ e quindi è il trasposto del gradiente della funzione $F^i(\mathbf{r})$; e quindi

$$J(\mathbf{r_0}) = \begin{bmatrix} \frac{\partial}{\partial x^1} F^1(\mathbf{r_0}) & \frac{\partial}{\partial x^2} F^1(\mathbf{r_0}) & \dots & \frac{\partial}{\partial x^n} F^1(\mathbf{r_0}) \\ \frac{\partial}{\partial x^1} F^2(\mathbf{r_0}) & \frac{\partial}{\partial x^2} F^2(\mathbf{r_0}) & \dots & \frac{\partial}{\partial x^n} F^2(\mathbf{r_0}) \\ \vdots & & & \\ \frac{\partial}{\partial x^1} F^m(\mathbf{r_0}) & \frac{\partial}{\partial x^2} F^m(\mathbf{r_0}) & \dots & \frac{\partial}{\partial x^n} F^m(\mathbf{r_0}) \end{bmatrix}$$

La matrice J si chiama la *matrice* *jacobiana* della funzione $\mathbf{F}(\mathbf{r})$, calcolata in $\mathbf{r}_0$. Quando $n = m$, il suo determinante si chiamo lo *jacobiano* della trasformazione.

Per mezzo della matrice jacobiana, possiamo scrivere

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}(\mathbf{r_0}) + J(\mathbf{r_0})(\mathbf{r} - \mathbf{r}_0) + \mathrm{o}(\,\|\mathbf{r} - \mathbf{r_0}\|\,)\,. \tag{4.9}$$

Quest'uguaglianza si chiama ancora *formula degli incrementi* *finiti*.

Ovviamente, lavorando per componenti, si potrà scrivere anche una formula di Taylor per funzioni da $\mathbf{R}^n$ in $\mathbf{R}^m$, su cui non indugiamo.

Come nel caso delle funzioni a valori in $\mathbf{R}$, la sola esistenza delle derivate parziali non implica né la continuità né la differenziabilità; ma la continuità delle derivate parziali in un insieme aperto implica la differenziabilità, su tale insieme, di ciascuna delle componenti $F^i(\mathbf{r})$ e quindi di $\mathbf{F}(\mathbf{r})$, ossia:

Teorema 149 *Le derivate parziali* $\frac{\partial}{\partial x^i}F^j(\mathbf{r_0})$ *per* $1 \le i \le n$, $1 \le j \le m$ *esistano* continue *in un intorno di* $\mathbf{r_0}$. *Allora, la funzione* $\mathbf{F}(\mathbf{r})$ *è differenziabile e quindi continua in* $\mathbf{r_0}$.

Rappresentiamo $\mathbf{R}^n$ rispetto alla base canonica e sia J una trasformazione lineare di $\mathbf{R}^n$ in sé. Col medesimo simbolo J indichiamo anche la matrice che rappresenta la trasformazione. Si sa che la trasformazione J trasforma un parallelepipedo in un altro e l'area del secondo è uguale all'area del primo moltiplicata per $|\det J|$.

Consideriamo ora una trasformazione differenziabile da $\mathbf{R}^2$ in sé:

$$x = x(u, v)\,, \qquad y = y(u, v)\,.$$

Fissiamo un punto $\mathbf{r_0} = (u_0, v_0)$ ed un rettangolo R con un vertice in $\mathbf{r_0}$. L'immagine di R è

$$\left\{ \begin{bmatrix} x(u,v) \\ y(u,v) \end{bmatrix}\,,\ (u,v) \in R \right\} = \left\{ \begin{bmatrix} x(u_0,v_0) \\ y(u_0,v_0) \end{bmatrix} + J \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix} + \mathbf{o}\left(\left\| \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix} \right\| \right) \right\}$$

ove J indica la matrice jacobiana della trasformazione. Quest'insieme non è un rettangolo. E' un insieme dalla struttura più complicata, ma approssimato dal parallelogramma

$$\tilde{R} = \left\{ \begin{bmatrix} x(u_0,v_0) \\ y(u_0,v_0) \end{bmatrix} + J \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix} \right\}\,.$$

Dal paragrafo 3.5

$$(\text{area di } \tilde{R}) = |\det J| \cdot (\text{area di } R).$$

Asserto analogo vale per le trasformazioni di $\mathbf{R}^3$.

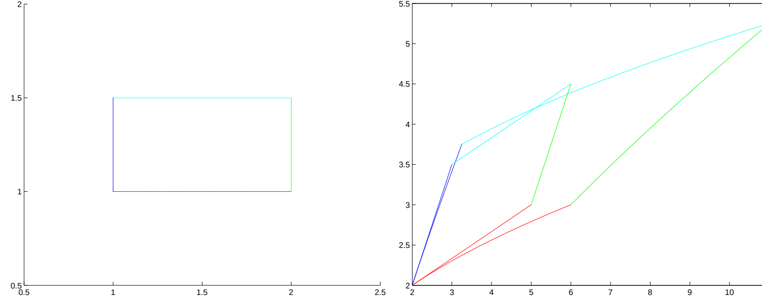
Quest'osservazione verrà utilizzata nel paragrafo 7.3.1.

Esempio 150 La figura 4.4 riporta, a sinistra, il rettangolo $R = [1,2] \times [1,3/2]$ del piano (u,v) e a destra la sua immagine mediante la trasformazione

$$x = u + (uv)^2\,, \qquad y = v^2 + uv\,.$$

Il parallelogramma a destra è $\tilde{R}$.

Figura 4.4: Un parallelogramma, la sua immagine e il parallelogramma che la approssima



Osservazione 151 Si vede dalla formula degli incrementi finiti che la matrice jacobiana ha, nel caso delle funzioni da $\mathbf{R}^n$ in $\mathbf{R}^m$, lo stesso ruolo che la derivata ha nel caso $n = m = 1$. Per questa ragione la matrice jacobiana talvolta si indica anche col simbolo

$$\frac{\partial \mathbf{F}(\mathbf{r})}{\partial \mathbf{r}} .$$

Per esempio, nel caso $n = m = 2$ e $\mathbf{F}(x, y) = \begin{bmatrix} f(x,y) & g(x,y) \end{bmatrix}'$, la matrice jacobiana si trova anche indicata con la notazione

$$\frac{\partial\left(f(x,y), g(x,y)\right)}{\partial(x,y)} . \qquad \blacksquare$$

Osservazione 152 (Sulle notazioni) Non esiste una notazione standard per lo jacobiano o per la matrice jacobiana. Nei testi di meccanica del continuo e di scienza delle costruzioni lo jacobiano viene indicato col simbolo ∇, come il gradiente. Si faccia attenzione però che se lo spazio d'arrivo ha dimensione 1 allora ∇="grad" indica un vettore colonna mentre ∇="matrice jacobiana" indica un vettore riga[2].

Si faccia anche attenzione a questo: nei testi di meccanica del continuo si considerano funzioni $\mathbf{u} = \mathbf{u}(\mathbf{X}, t)$ con $\mathbf{X}$ e $\mathbf{u}$ in $\mathbf{R}^3$ e $t \in \mathbf{R}$. La notazione $\nabla \mathbf{u} = \nabla \mathbf{u}(\mathbf{X}, t)$ indica la matrice jacobiana fatta rispetto al vettore $\mathbf{X}$, ossia

$$\nabla \mathbf{u}(\mathbf{X}, t) = \begin{bmatrix} \frac{\partial u_1(X_1,X_2,X_3,t)}{\partial X_1} & \frac{\partial u_1(X_1,X_2,X_3,t)}{\partial X_2} & \frac{\partial u_1(X_1,X_2,X_3,t)}{\partial X_3} \\ \frac{\partial u_2(X_1,X_2,X_3,t)}{\partial X_1} & \frac{\partial u_2(X_1,X_2,X_3,t)}{\partial X_2} & \frac{\partial u_2(X_1,X_2,X_3,t)}{\partial X_3} \\ \frac{\partial u_3(X_1,X_2,X_3,t)}{\partial X_1} & \frac{\partial u_3(X_1,X_2,X_3,t)}{\partial X_2} & \frac{\partial u_3(X_1,X_2,X_3,t)}{\partial X_3} \end{bmatrix} .$$

I testi di meccanica del continuo usano una notazione interessante per le derivate: le derivate si indicano con indici numerici preceduti da virgola. Per esempio, sia

$$\mathbf{u}(x_1, x_2, x_3) = \begin{bmatrix} u_1(x_1,x_2,x_3) & u_2(x_1,x_2,x_3) & u_3(x_1,x_2,x_3) \end{bmatrix}' .$$

---

[2]va notato che in certi testi di meccanica del continuo, il gradiente di una funzione a valori in $\mathbf{R}$ si intende essere un vettore riga; ossia in tali libri si confonde il gradiente col vettore che rappresenta il differenziale.

Il simbolo
$$\mathbf{u}_{i,j}$$
indica la derivata rispetto ad x_j della componente u_i; ossia il numero prima della virgola specifica la componente e quello dopo specifica la variabile rispetto a cui si deriva. Questo simbolo si può iterare e per esempio
$$\mathbf{u}_{2,1\,3} = \frac{\partial^2}{\partial x_1 \partial x_3} u_2(x_1, x_2, x_3)\,. \quad \blacksquare$$

4.5.1 Regole di calcolo della matrice jacobiana

Usiamo il simbolo $J_{\mathbf{F}}(\mathbf{r})$ per indicare la matrice jacobiana della funzione $\mathbf{F}$ nel punto $\mathbf{r}$. E' immediatamente evidente che
$$J_{A\mathbf{F}}(\mathbf{r}) = AJ_{\mathbf{F}}(\mathbf{r}) \qquad \text{per ogni matrice costante } A;$$
$$J_{\mathbf{F}+\mathbf{G}}(\mathbf{r}) = J_{\mathbf{F}}(\mathbf{r}) + J_{\mathbf{G}}(\mathbf{r})\,.$$

Supponiamo ora di esprimere $\mathbf{r} \in \mathbf{R}^n$ come funzione
$$\mathbf{r} = \mathbf{r}(\mathbf{u})\,, \qquad \mathbf{u} \in \mathbf{R}^k\,.$$

Vale

Teorema 153 (della derivazione a catena) *Sia $\mathbf{F}(\mathbf{r})$ di classe C^1 in una regione Ω di $\mathbf{R}^n$; $\mathbf{r}(\mathbf{u})$ sia di classe C^1 in una regione $\tilde{\Omega} \subseteq \mathbf{R}^k$, ed a valori in Ω. Sia $\mathbf{G}(\mathbf{u}) = \mathbf{F}(\mathbf{r}(\mathbf{u}))$. Si ha*
$$J_{\mathbf{G}}(\mathbf{u_0}) = J_{\mathbf{F}}(\mathbf{r}(\mathbf{u_0}))J_{\mathbf{r}}(\mathbf{u_0})\,.$$

Il prodotto tra matrici si intende come prodotto righe per colonne.

Omettiamo la dimostrazione.

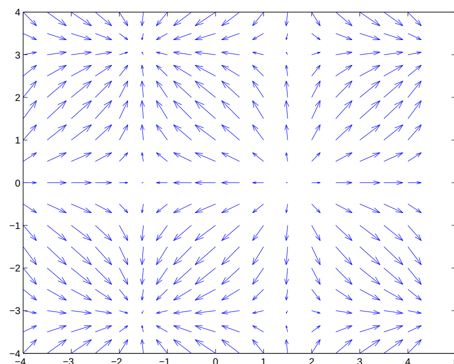
4.6 Campi vettoriali

Nel trattare le funzioni a valori vettoriali, abbiamo sempre specificato "a valori in $\mathbf{R}^m$". Si ricordi che $\mathbf{R}^m$ è uno spazio di vettori liberi, equivalentemente, applicati nell'origine. In molte applicazioni si devono considerare funzioni che associano ad un punto $\mathbf{r}$ di $\mathbf{R}^n$ un vettore ad n dimensioni, applicato nel punto stesso. Ossia si devono studiare trasformazioni che trasformano $\mathbf{r}$ nella coppia $(\mathbf{r}, \mathbf{V}(\mathbf{r}))$. Trasformazioni di questo tipo si chiamano *campi vettoriali*. E' comune usare la lettera $\mathbf{V}$ per indicare campi vettoriali. Si scrive cioè $\mathbf{V}(\mathbf{r})$ sottintendendo che questo vettore è applicato in $\mathbf{r}$, invece di scrivere la notazione completa, ma pedante, $(\mathbf{r}, \mathbf{V}(\mathbf{r}))$

La figura 4.5 illustra il modo con cui usualmente si rappresentano i campi vettoriali: da ciascun punto si fa uscire una freccia che indica direzione e verso del vettore. La lunghezza della freccia è proporzionale al modulo.

Quando si lavora con campi vettoriali, i concetti di continuità e di differenziabilità sono concettualmente diversi da quelli incontrati per i vettori liberi, perché non si fanno operazioni tra vettori applicati in punti diversi. Per questo la definizione di continuità di un campo

Figura 4.5: un campo vettoriale



vettoriale si definisce confrontando il vettore $\mathbf{V}(\mathbf{r})$, applicato in $\mathbf{r}$ col vettore PARALLELO a $\mathbf{V}(\mathbf{r}')$, anch'esso applicato in $\mathbf{r}$ invece che in $\mathbf{r}'$.

Fortunatamente, la somma di vettori applicati nel medesimo punto si fa operando per componenti, e queste non mutano per spostamenti per parallelismo. Si ha quindi che un campo vettoriale è continuo se e solo se sono continue le sue componenti.

Discorso analogo vale per le derivate. Si fissi $\mathbf{r_0}$ ed un vettore libero $\mathbf{v}$. Si considerino i vettori

$$\mathbf{r_0} + t\mathbf{v}$$

che sono i vettori di una retta parallela a $\mathbf{v}$, passante per $\mathbf{r_0}$. Si chiama *derivata secondo il* *vettore* $\mathbf{v}$ il limite

$$D_{\mathbf{v}}\mathbf{V}(\mathbf{r_0}) = \lim_{t\to 0} \frac{\mathbf{V}(\mathbf{r_0} + t\mathbf{v}) - \mathbf{V}(\mathbf{r_0})}{t} .$$

La differenza al numeratore si calcola supponendo di traslare per parallelismo il vettore $\mathbf{V}(\mathbf{r_0} + t\mathbf{v})$ applicandolo nel punto $\mathbf{r_0}$. Dunque, anche il vettore $D_{\mathbf{v}}\mathbf{V}(\mathbf{r_0})$ è applicato in $\mathbf{r_0}$.

Se $\mathbf{v}$ è il versore $\mathbf{e}_i$, tale derivata si chiama anche *derivata* *direzionale*.

Il teorema di derivazione della funzione composta immediatamente dà:

Teorema 154 *Sia* $\mathbf{V}(\mathbf{r})$ *un campo vettoriale le cui componenti sono funzioni di classe* C^1*. Esso ammette derivate direzionali in tutte le direzioni, e vale*

$$D_{\mathbf{v}}\mathbf{V}(\mathbf{r_0}) = J_{\mathbf{V}}(\mathbf{r_0})\mathbf{v} \tag{4.10}$$

In questo teorema, il simbolo $J_{\mathbf{V}}(\mathbf{r_0})$ indica la matrice jacobiana della trasformazione che al vettore libero $\mathbf{r}$ associa il vettore libero $\mathbf{V}(\mathbf{r})$. Il vettore $D_{\mathbf{v}}\mathbf{V}(\mathbf{r_0})$ si intende applicato in $\mathbf{r_0}$.

4.6.1 Operatori differenziali e campi vettoriali

Abbiamo visto che il gradiente di una funzione a valori scalari è il vettore

$$\nabla f(\mathbf{r}) = \begin{bmatrix} f_{x^1}(\mathbf{r}) \\ f_{x^2}(\mathbf{r}) \\ \vdots \\ f_{x^n}(\mathbf{r}) \end{bmatrix} .$$

Per molte applicazioni, conviene considerare questo un vettore applicato nel punto $\mathbf{r}$. Dunque il gradiente associa un campo vettoriale ad una funzione a valori scalari.

Si noti che

$$[\nabla_{\mathbf{u}} f(\mathbf{r}(\mathbf{u}))]' = [\nabla_{\mathbf{r}} f(\mathbf{r}(\mathbf{u}))]' \, J_{\mathbf{r}}[\mathbf{u}]$$

ove l'apice indica la trasposizione e l'indice sotto il simbolo ∇ indica le variabili rispetto alle quali si calcola il gradiente.

Introduciamo una notazione comoda: sia $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ la base canonica di $\mathbf{R}^n$. Scriviamo in modo formale

$$\begin{aligned} \nabla &= \mathbf{e}_1 \frac{\partial}{\partial x^1} + \mathbf{e}_2 \frac{\partial}{\partial x^2} + \cdots + \mathbf{e}_n \frac{\partial}{\partial x^n} \\ &= \mathbf{e}_1 \partial_1 + \mathbf{e}_2 \partial_2 + \cdots + \mathbf{e}_n \partial_n \\ &= \begin{bmatrix} \partial_1 \\ \partial_2 \\ \vdots \\ \partial_n \end{bmatrix} . \end{aligned}$$

Allora il gradiente si ottiene formalmente "moltiplicando" il "vettore" ∇ per lo "scalare" $f(\mathbf{r})$, e eseguendo le operazioni di derivazione.

Introdotto il simbolo ∇, viene naturale introdurre le operazioni

$$\nabla \cdot \mathbf{V}(\mathbf{r}) \,, \qquad \nabla \wedge \mathbf{V}(\mathbf{r})$$

operando formalmente come se si trattasse di prodotti scalari o vettoriali (e quindi la seconda si definisce solo in $\mathbf{R}^3$). La prima dà la *divergenza* del campo vettoriale:

$$\mathrm{div}\mathbf{V}(\mathbf{r}) = \frac{\partial}{\partial x^1} V^1(\mathbf{r}) + \frac{\partial}{\partial x^2} V^2(\mathbf{r}) + \cdots + \frac{\partial}{\partial x^n} V^n(\mathbf{r}) \,.$$

Il risultato della seconda operazione, DEFINITA SOLO IN $\mathbf{R}^3$, si chiama *rotore*: il rotore del campo vettoriale

$$\mathbf{V}(x,y,z) = u(x,y,z)\mathbf{i} + v(x,y,z)\mathbf{j} + w(x,y,z)\mathbf{k}$$

è quindi (sottintendendo le variabili indipendenti) il campo vettoriale

$$\mathrm{rot}\mathbf{V} = \nabla \wedge \mathbf{V} = (w_y - v_z)\mathbf{i} + (u_z - w_x)\mathbf{j} + (v_x - u_y)\mathbf{k} \,.$$

Si noti che, sviluppando formalmente secondo gli elementi della prima riga,

$$\mathrm{rot}\mathbf{V} = \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u & v & w \end{bmatrix} .$$

Sia ora $u(x, y, z)$ una funzione due volte derivabile. Calcoliamo prima il gradiente di $u(x, y, z)$ e poi la divergenza del gradiente:

$$\nabla \cdot (\nabla u(x, y, z)) = u_{xx}(x, y, z) + u_{yy}(x, y, z) + u_{zz}(x, y, z) .$$

Usa definire Δ, che si chiama *laplaciano*, ponendo

$$\Delta u(x, y, z) = u_{xx}(x, y, z) + u_{yy}(x, y, z) + u_{zz}(x, y, z) .$$

Formalmente si scrive

$$\Delta = \nabla \cdot \nabla .$$

Si vede facilmente che queste considerazioni valgono per funzioni di qualunque numero di variabili.

Osservazione 155 (Sulle notazioni) Proseguiamo quanto si è detto all'osservazione 152. Facendo intervenire le notazioni dei campi vettoriali, i testi di meccanica del continuo e di scienza delle costruzioni scrivono la prima formula degli incrementi finiti in questo modo. Sia $\mathbf{u}(\mathbf{x})$ una funzione di tre variabili a valori in $\mathbf{R}^3$. Usando la notazione "con la virgola" introdotta all'osservazione 152, la prima formula degli incrementi finiti è

$$\mathbf{u}(\mathbf{x}) = \mathbf{u}(\mathbf{0}) + \begin{bmatrix} u_{1,1}(\mathbf{0}) & u_{1,2}(\mathbf{0}) & u_{1,3}(\mathbf{0}) \\ u_{2,1}(\mathbf{0}) & u_{2,2}(\mathbf{0}) & u_{2,3}(\mathbf{0}) \\ u_{3,1}(\mathbf{0}) & u_{3,2}(\mathbf{0}) & u_{3,3}(\mathbf{0}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \mathrm{o}(||\mathbf{x}||) .$$

Invece di scrivere in questo modo, viene scritto:

$$\mathbf{u}(\mathbf{x}) = \mathbf{u}(\mathbf{0}) + (\mathbf{x} \cdot \nabla)\, \mathbf{u}(\mathbf{0}) + \mathrm{o}(||\mathbf{x}||)$$

ove $(\mathbf{x} \cdot \nabla)$ indica un "prodotto scalare formale" tre i due vettori colonna $\mathbf{x}$ e ∇,

$$\mathbf{x} \cdot \nabla = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} \partial_1 \\ \partial_2 \\ \partial_3 \end{bmatrix} = x_1\partial_1 + x_2\partial_2 + x_3\partial_3 .$$

Si trova quindi un operatore differenziale che si applica al campo vettoriale $\mathbf{u}(\mathbf{x})$, applicandolo a ciascuna delle sue componenti:

$$\begin{aligned} (\mathbf{x} \cdot \nabla)\, \mathbf{u}(\mathbf{x}) &= \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} \partial_1 \\ \partial_2 \\ \partial_3 \end{bmatrix} \right) \mathbf{u}(\mathbf{x}) \\ &= (x_1\partial_1 + x_2\partial_2 + x_3\partial_3)\, \mathbf{u}(\mathbf{x}) = \begin{bmatrix} x_1 u_{1,1} + x_2 u_{1,2} + x_3 u_{1,3} \\ x_1 u_{2,1} + x_2 u_{2,2} + x_3 u_{2,3} \\ x_1 u_{3,1} + x_2 u_{3,2} + x_3 u_{3,3} \end{bmatrix} \end{aligned}$$

proprio come si voleva ottenere. ∎

4.7 Appendici

4.7.1 Appendice: Rappresentazione di funzioni di due variabile

In quest'appendice mostriamo alcune funzioni di due variabili a valori reali, ed il modo di rappresentarle. Sostanzialmente, ci sono tre modi. Descriviamoli prima di tutto in astratto e poi illustriamoli su opportuni esempi.

Sia $f(x, y)$ una funzione. Il modo più ovvio di rappresentarla consiste nel costruirne il grafico, ossia nel costruire l'insieme

$$\{(x, y, z) \mid z = f(x, y)\} \subseteq \mathbf{R}^3 .$$

Naturalmente questo richiede la costruzione di un modello per esempio di gesso o di plastica. In pratica rappresenteremo il grafico su un foglio, mediante opportune tecniche di disegno, che diano l'illusione della profondità. Inoltre, talvolta conviene rappresentare, invece della superficie, solamente una famiglia di linee sulla superficie.

Il secondo metodo consiste nel tagliare il grafico a quota c, e quindi nel proiettare l'insieme intersezione sul piano (x, y). Si disegna cioè *sul piano* (x, y), l'insieme delle soluzioni dell'equazione

$$f(x, y) = c .$$

Quest'insieme si chiama *insieme di livello*. In molti casi[3] un insieme di livello è una "curva" e quindi si parla di *curva di livello*. Disegnando "numerose" curve di livello si può avere un'informazione sul comportamento della funzione: per esempio, le curve di livello sono più fitte dove il grafico della funzione è più ripido.

Il terzo metodo consiste nel colorare il piano (x, y) colorandolo con colori "freddi", per esempio blu, dove i valori della funzione sono più piccoli e con colori caldi, per esempio rosso, dove la funzione prende valori maggiori.

Molto spesso conviene combinare questi metodi, attribuendo i colori alla superficie o alle sue curve di livello.

Vediamo ora un esempio. Consideriamo la funzione seguente, il cui grafico è un paraboloide di rotazione:

$$f(x, y) = x^2 + y^2 .$$

La figura 4.6 mostra vari modi di rappresentare il grafico di questa funzione. La figura 4.7 mostra a sinistra la funzione rappresentata mediante le sue curve di livello (e quindi mediante una rappresentazione sul piano (x, y)) ed a destra il grafico (fatto con fili) della medesima funzione, sovrapposto alle curve di livello.

Ciò detto, passiamo ad esaminare alcuni esempi di funzioni.

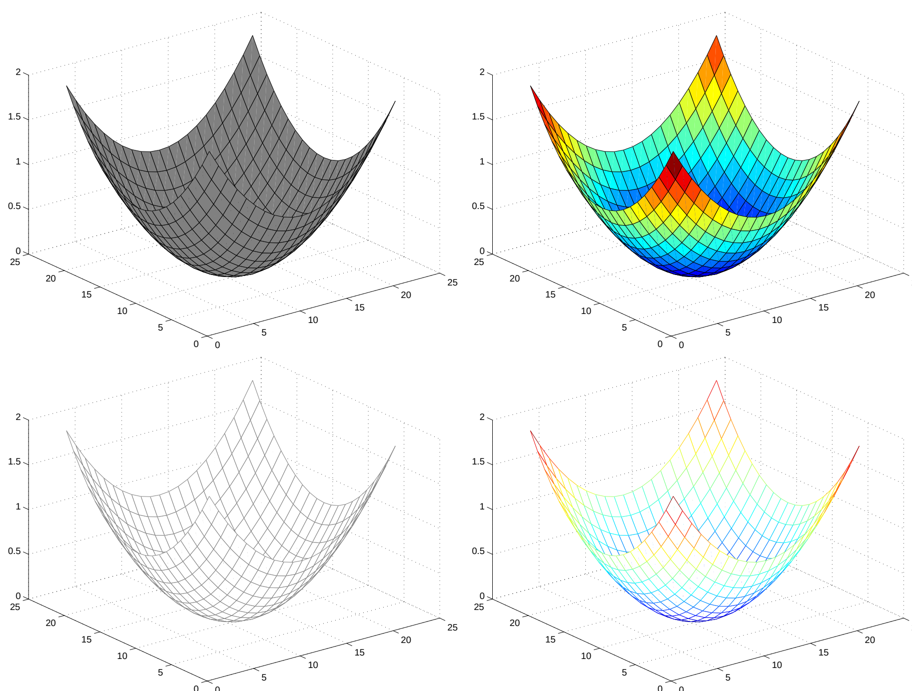
Funzioni costanti rispetto ad una variabile Può accadere che una funzione $f(x, y)$ sia "costante rispetto ad y"; ossia che valga

$$f(x, y) = f(x, y')$$

per ogni scelta di y e di y', purché la x prenda il medesimo valore a destra ed a sinistra.

[3] come proveremo al Capitolo 5

Figura 4.6: Paraboloide di rotazione



Vediamo le conseguenze “grafiche” di questa proprietà. La funzione è costante sulle rette verticali del piano (x, y), che pertanto sono tutte contenute in insiemi di livello. Per esempio la funzione

$$f(x, y) = \sin x \tag{4.11}$$

ha gli insiemi di livello rappresentati nella figura 4.8 a sinistra. Rette del medesimo colore corrispondono al medesimo livello e quindi gli insiemi di livello in quest'esempio non sono “curve”, ma sono insiemi di (infinite) rette verticali.

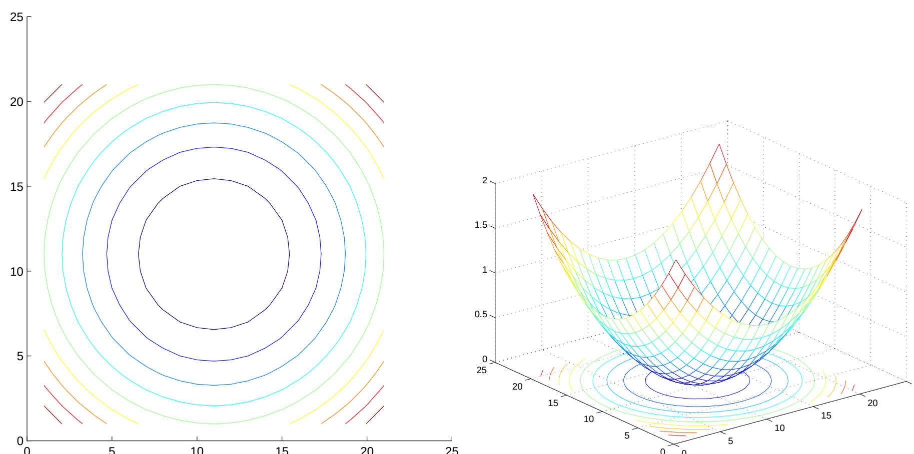
Dal punto di vista del grafico, il grafico della funzione è unione di rette orizzontali, e quindi tutte parallele tra loro.

Il termine *superficie* verrà introdotto più avanti, ma è un fatto che sono “superfici” i grafici di funzioni (differenziabili). Una superficie che è unione di rette tutte tra loro parallele si chiama *cilindro*. Dunque, i grafici di funzioni $f(x, y)$ costanti rispetto ad una variabile sono cilindri. Il cilindro corrispondente alla funzione in esame è nella figura (4.8), a destra.

Funzioni omogenee Si chiamano *funzioni omogenee* di grado ν (non necessariamente intero) quelle funzioni che verificano l'uguaglianza

$$f(rx, ry) = |r|^{\nu} f(x, y) .$$

Figura 4.7: Paraboloide di rotazione e sue curve di livello



Se $\nu = 0$ la funzione è costante sulle rette che escono dall'origine, che pertanto appartengono ad insiemi di livello. Un esempio è la funzione

$$f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}\,.$$

Dato che il denominatore si annulla in $(0, 0)$, per evitare difficoltà di tipo numerico, consideriamola nel quadrati $0,1 \leq x \leq 1$, $0,1 \leq y \leq 1$. La figura 4.9, a sinistra, presenta le curve di livello della funzione. Come si è detto, queste sono rette, ma problemi di tipo numerico vicino a $(0, 0)$ (dove si annullano sia il numeratore che il denominatore) provocano le distorsioni visibili in figura. Il valore della funzione varia da retta a retta, ma rimane costante sulle singole rette, come mostra la figura 4.9, a destra. In questa figura abbiamo rappresentato, in verde, anche il piano $z = 0$.

Se la funzione è omogenea di grado 1 allora moltiplicando per r sia x che y, il valore di z viene anch'esso moltiplicato per r. Dunque, se un punto (x_0, y_0, z_0) è sul grafico, anche i punti (rx_0, ry_0, rz_0) con $r > 0$ sono sul grafico. Al variare di r, questi descrivono semirette passanti per l'origine: il grafico è unione di semirette passanti per l'origine. Un insieme unione di rette per un medesimo punto si chiama *cono*. E quindi il grafico di una funzione omogenea di grado 1 è un semicono (con vertice nell'origine).

La figura 4.10, a sinistra, illustra la situazione nel caso della funzione

$$f(x, y) = \frac{xy}{x + y}\,.$$

La figura riporta sia il grafico della funzione sia alcune rette che escono dall'origine e che appartengono al grafico.

La figura 4.10, a destra, riporta invece il grafico della funzione

$$f(x, y) = \frac{\sqrt{xy}}{x + y}\,,$$

Figura 4.8: La funzione (4.11)

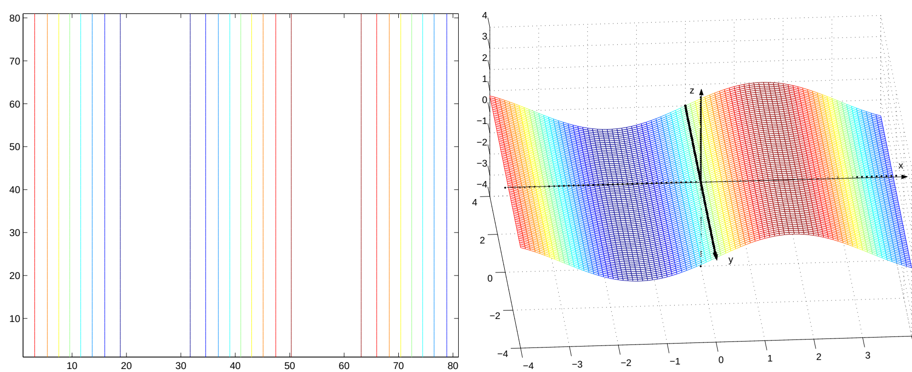
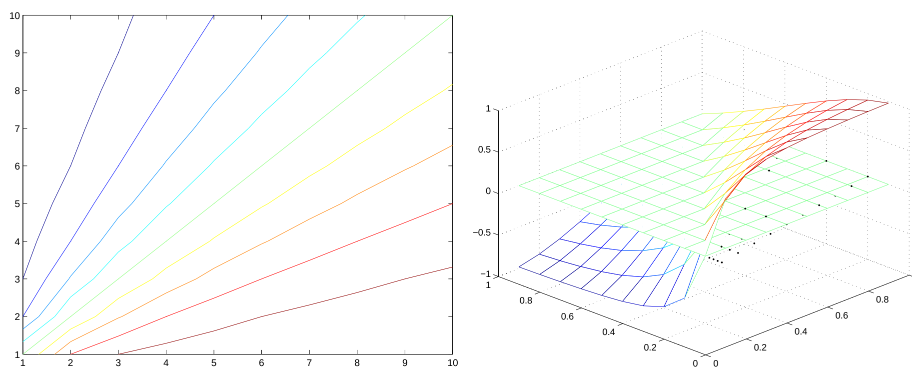


Figura 4.9: Una funzione omogenea



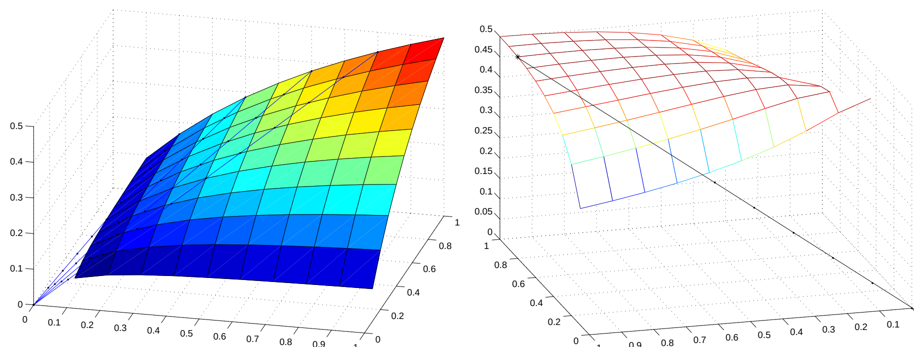
omogenea di grado 1/2. La figura riporta anche una retta per l'origine e per un punto del grafico che, come si vede, non giace sul grafico.

Simmetrie del grafico, funzioni radiali e superfici di rotazione Il grafico di una funzione di più variabili può avere proprietà di simmetria, che estendono quelle note per le funzioni di una variabile. Per esempio, diremo che una funzione $f(x, y)$ è

- *dispari* se $f(x, y) = -f(-x, -y)$;
- *pari* se $f(x, y) = f(-x, -y)$.

Come nel caso delle funzioni di una variabile, queste proprietà corrispondono a certe simmetrie del grafico rispetto all'origine. Naturalmente, si possono avere altre possibilità che non si incontrano nel caso di funzioni di una variabile (per esempio, simmetrie rispetto ad una retta per l'origine, magari ad uno degli assi coordinati).

Figura 4.10: Funzione omogenea di grado 1/2



A noi ora interessa un caso molto particolare di funzione pari di due variabili, ottenuta a partire da funzioni di una sola variabile.

Si consideri la funzione $t \mapsto f(t)$ di **una sola variabile** definita su un intervallo[4] $[a, b]$ e a questa si associ la funzione di **due variabili**

$$(x, y) \mapsto f\left(\sqrt{x^2 + y^2}\right)$$

ossia la funzione

$$z = f\left(\sqrt{x^2 + y^2}\right) , \quad \operatorname{dom} f = \left\{(x, y) \, , \ \sqrt{x^2 + y^2} \in [a, b]\right\} .$$

I grafici di queste funzioni hanno una proprietà particolare. Sia

$$(x_0, y_0) \in \operatorname{dom} f\left(\sqrt{x^2 + y^2}\right) \quad \text{ed} \quad r_0 = \sqrt{x_0^2 + y_0^2} .$$

Il numero r_0 è la distanza di (x_0, y_0) dall'origine. Qualunque altro punto (x, y) che dista r_0 dall'origine, ossia per cui

$$\sqrt{x^2 + y^2} = r_0$$

appartiene al dominio della funzione ed inoltre,

$$f\left(\sqrt{x^2 + y^2}\right) = f\left(\sqrt{x_0^2 + y_0^2}\right) .$$

Dunque il dominio della funzione è una corona circolare di centro l'origine (un disco se $a = 0$, l'esterno di un disco se $a > 0$ ed $r = +\infty$), e tutte le circonferenze di centro l'origine sono curve di livello della funzione.

Questo si interpreta in questo modo: **se il dominio della funzione contiene un punto P, esso contiene tutta la circonferenza di centro O che passa per P. Inoltre su tale circonferenza la funzione è costante.** Ossia **se il punto P ruota su**

[4]facciamo il caso di un intervallo chiuso, ma quello che andiamo a dire si adatta in ogni caso, anche nel caso che si tratti di una semiretta.

una circonferenza di centro O il corrispondente punto del grafico ruota su una circonferenza parallela al piano x, y, centrata sull'asse delle quote.

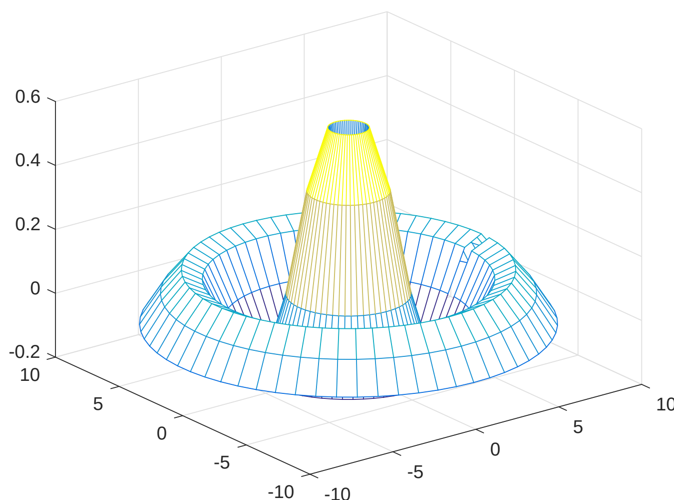
Detto in altro modo, **ogni piano perpendicolare all'asse delle quote se interseca il grafico della funzione lo interseca lungo una circonferenza (oppure in un solo punto).**

Una superficie come quella ora descritta si chiama *superficie di rotazione.*

L'intersezione tra un piano $z = z_0$ e il grafico, se non è vuota, si chiama un *parallelo* della superficie mentre l'intersezione tra il grafico ed ogni piano passante per l'asse delle quote si chiama *meridiano* della superficie di rotazione: **la superficie di rotazione può anche pensarsi ottenuta ruotando un meridiano intorno all'asse delle quote.**

Il grafico nella figura 4.11 mostra una superficie di rotazione, con indicati alcuni paralleli ed alcuni meridiani.

Figura 4.11: Superficie di rotazione



Osservazione 156

- Una funzione $g(x, y) = f\left(\sqrt{x^2 + y^2}\right)$ che definisce una superficie di rotazione si chiama anche una *funzione radiale.* Più in generale, si chiama *funzione radiale* ogni funzione su $\mathbf{R}^n$ il cui valore in ogni $\mathbf{x}$ del suo dominio dipende solo da $\|\mathbf{x}\|$. Tale funzione si esprime come $f(\|x\|)$ con f funzione di una variabile reale.
- La funzione $g\left(x^2 + y^2\right)$ è una funzione radiale, anche se nella sua espressione la radice non compare. Infatti,

$$g\left(x^2 + y^2\right) = g\left(\left(\sqrt{x^2 + y^2}\right)^2\right) . \quad \blacksquare$$

Funzioni periodiche Una funzione di due variabili, $f(x,y)$, si dice periodica quando esiste un vettore $\mathbf{v_0} = (\xi_0, \eta_0)$ tale che

$$f(\mathbf{x}) = f(x+\xi_0, y+\eta_0) = f(x,y)$$

per ogni (x,y) del suo dominio. Un esempio è la funzione

$$f(x,y) = (\sin x)(\sin 2 * y)\,.$$

Se

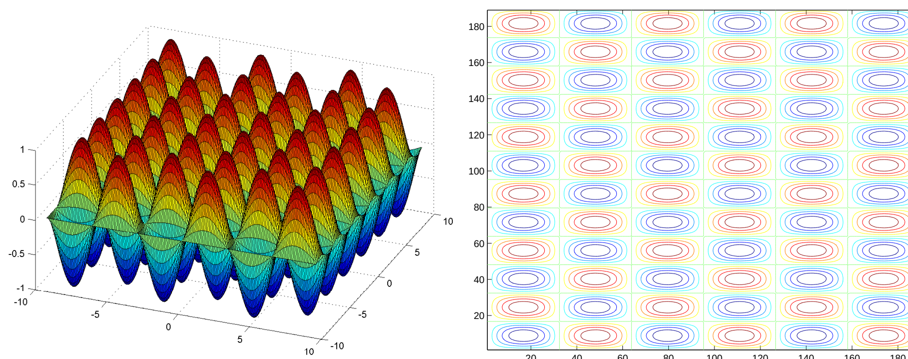
$$\mathbf{v} = 2\pi\mathbf{i} + \pi\mathbf{j}$$

si ha

$$f(\mathbf{x}+\mathbf{v}) = f(x+2\pi, y+\pi) = [\sin(x+2pi)]\,[\sin(2(y+\pi))] = f(x,y)\,.$$

La figura 4.12 mette a confronto il grafico della funzione $f(x,y)$, a sinistra, con le sue curve di livello, a destra.

Figura 4.12: Una funzione periodica



4.7.2 Appendice: Propagazione ondosa

In molte applicazioni si incontrano funzioni della forma

$$f(\phi(\mathbf{x}) - vt) \qquad \text{oppure} \qquad f(\phi(\mathbf{x}) + vt)$$

con $v > 0$.

In queste applicazioni la variabile t si interpreta come "tempo" mentre la variabile $\mathbf{x}$ indica una posizione nello spazio. Dunque la funzione $f(\phi(\mathbf{x}) \pm vt)$ descrive una configurazione $f(\phi(\mathbf{x}))$ che "si muove" al passare del tempo e ciò suggerisce un linguaggio particolare, che illustriamo con riferimento al caso $\mathbf{x} = x \in \mathbf{R}$ e $\phi(x) = x$ oppure $\phi(x) = mx$.

In queste applicazioni, la funzione $f(x)$ si chiama *onda*. Potremmo pensare al suo grafico come alla configurazione all'istante $t = 0$ di un corpo elastico. Vediamo come si interpetano le due funzioni

$$F_1(x,t) = f(x - vt)\,, \qquad F_2(x,t) = f(x + vt)$$

con $v > 0$.

Consideriamo la funzione $F_1(x,t)$. Fissiamo un valore t_0. La funzione $x \to f(x - vt_0)$ si può interpretare in due modi diversi: come la funzione $f(x)$ rappresentata rispetto ad un sistema di riferimento ottenuto traslando l'origine nella posizione $-vt_0$ dell'asse x. Dunque in dietro, perché $v > 0$. Questo modo di intendere le cose è il più comune nelle applicazioni geometriche. Invece, nelle applicazioni fisiche e nell'analisi si incontra più frequentemente un'interpretazione diversa: si pensa ad $f(x)$ ed $f(x - vt_0)$ come a due diverse funzioni rappresentate rispetto al medesimo sistema di riferimento. In tal caso, il grafico della funzione $f(x - vt_0)$ è ottenuto traslando in avanti il grafico di $f(x)$.

Se t si interpreta come "tempo", nell'unità di tempo il grafico è andato avanti di una quantità v. Dunque, v rappresenta la velocità dello spostamento del grafico. Pensando al grafico di $f(x)$ come ad un'"onda", si dice che v è la *velocità di propagazione dell'onda.*

Le stesse considerazioni si possono ripetere per la funzione $F_2(x,t)$ con la sola differenza che ora il grafico si sposta in dietro (grazie al fatto che v si è scelto positivo).

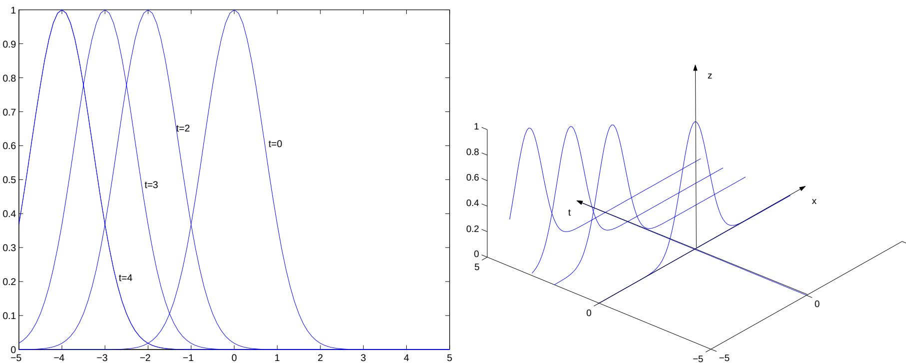
Le considerazioni precedenti suggeriscono di chiamare la funzione $F_1(x,t) = f(x - vt)$ *onda progressiva* mentre la funzione $F_1(x,t) = f(x + vt)$ si chiama *onda regressiva* (ricordiamo che $v > 0$).

La figura 4.13 riporta i grafici delle funzioni

$$e^{-(x+t)^2}\,. \tag{4.12}$$

a sinistra sono riportati i grafici delle funzioni $x \to e^{-(x+t)^2}$ nel medesimo sistema di assi cartesiani (x,y), per diversi valori di t. A destra è stato costruito il grafico della funzione di due variabili $(x,t) \to e^{-(x+t)^2}$, nello spazio riferito agli assi cartesiani (x,t,z). Questo grafico è stato sezionato con i piani $t = 0$, $t = 2$, $t = 3$ e $t = 4$, e sono stati disegnati i grafici risultanti da queste sezioni.

Figura 4.13: La funzione (4.12)

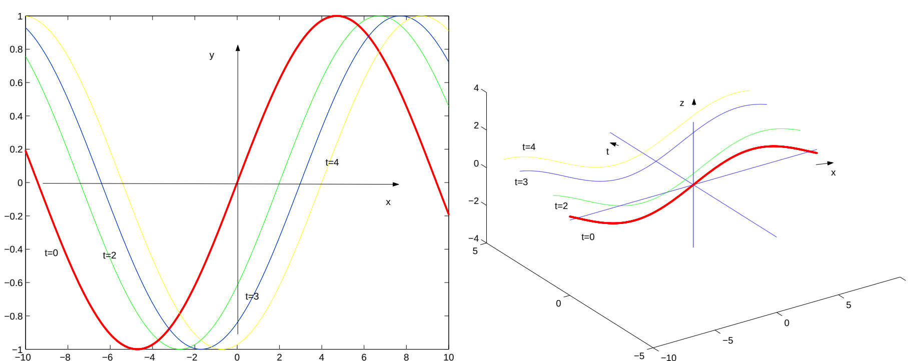


Il termine “onda” o “moto ondoso” usato per descrivere la situazione illustrata sopra è particolarmente intuitivo nel caso di funzioni $f(x)$ periodiche. La figura 4.14 riporta grafici analoghi a quelli della figura 4.13, ma questa volta con la funzione

$$f(x) = \sin(mx) \ . \tag{4.13}$$

Si noti che questa funzione è periodica di periodo $2\pi/m$. Il valore scelto nella figura per il periodo è 3.

Figura 4.14: La funzione (4.13)



Il caso in cui $f(x) = \sin(mx)$ oppure $f(x) = \cos(mx)$ è particolarmente importante nelle applicazioni. Le “onde” costruite a partire da queste funzioni si dicono *onde armoniche* e in questo caso si usano notazioni e definizioni peculiari che verranno illustrate al paragrafo 4.7.2.

Consideriamo ora la funzione $F_1(x,t) = f(x - vt)$. Questa funzione prende il valore $f(k)$ in tutti i punti (x,t) nei quali

$$x - vt = k\,.$$

Per ogni k si viene così a descrivere una retta nei punti della quale l'onda ha il medesimo valore. Ciò suggerisce di dare un nome alla quantità $x - vt$: questa si chiama la *fase* dell'onda progressiva. Al trascorrere del tempo, si rimane sulla stessa fase k se si passa dal punto x al punto $x + vt$, ossia se si sposta il punto osservato x in avanti con velocità v. Per questa ragione, il numero v si chiama anche (e più frequentemente) *velocità di fase*.

Considerazioni analoghe per le onde regressive si lasciano per esercizio.

Si noti che talvolta un'onda viene scritta come

$$f(hx - mt) = f\left(h(x - (m/h)t)\right) \ .$$

In tal caso, la velocità di fase è $v = m/h$.

Se $\mathbf{x} \in \mathbf{R}^2$ oppure $\mathbf{x} \in \mathbf{R}^3$ si incontra un nuovo concetto, che illustriamo con riferimento a $\mathbf{x} \in \mathbf{R}^2$, $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}'$. Consideriamo le due onde progressive seguenti:

$$\Phi_1(\mathbf{x},t) = \Phi_1(x,y,t) = f(ax + by - vt)\,, \qquad \Phi_2(\mathbf{x},t) = \Phi_2(x,y,t) = f\left(x^2 + y^2 - vt\right) \ .$$

Ci si può chiedere su quali insiemi del piano (x, y) l'onda ha la medesima fase per ogni fissato valore di t. Lavorando con la funzione $\Phi(\mathbf{x}, t)$, fissato $t = t_0$, la fase vale k nei punti della retta

$$ax + by = k + vt_0\,;$$

nel caso della funzione $\Phi_2(\mathbf{x}, t)$ fissato $t = t_0$, la fase vale k nei punti della circonferenza

$$x^2 + y^2 = k + vt_0\,.$$

Per questa ragione, l'onda $\Phi_1(\mathbf{x}, t)$ si chiama *onda piana* mentre l'onda $\Phi_2(\mathbf{x}, t)$ si chiama *onda sferica*.

Onde e moto armonico

Ricordiamo che si chiama *moto armonico* il moto di un punto lungo una circonferenza, se questo avviene con velocità angolare costante.

I punti di una circonferenza di raggio A hanno coordinate

$$(A\cos(\omega t + \phi)\,, A\sin(\omega t + \phi))\ . \tag{4.14}$$

Il numero $(2\pi)/\omega$ è il tempo necessario a descrivere una volta tutta la circonferenza e quindi

$$T = \frac{2\pi}{\omega}$$

è il *periodo* del moto mentre il suo reciproco

$$f = \frac{\omega}{2\pi}$$

si chiama *frequenza* del moto. La frequenza indica quanti giri (o parti di giro) percorre il corpo lungo la circonferenza in un'unità di tempo.

Il numero di radianti di cui varia la posizione del corpo nell'unità di tempo è ω e per questo il coefficiente ω si chiama *velocità angolare*. Dunque,

$$\omega = 2\pi f\,.$$

Dato che $\omega = (2\pi)/T$, la (4.14) si può anche scrivere (si confronti con l'osservazione 82 dove però il periodo è 2T)

$$\left(A\cos\frac{2\pi}{T}t\,, A\sin\frac{2\pi}{T}t\right)\ .$$

Questo legame col moto armonico suggerisce di chiamare *onde armoniche* le onde

$$A\cos(kx - \omega t + \phi)\,, \qquad A\sin(kx - \omega t + \phi)$$

(e le analoghe onde regressive, il cui esame si lascia per esercizio).

Consideriamo per fissare le idee l'onda progressiva

$$A\sin(kx - \omega t + \phi)\,.$$

Il numero A si chiama l' *ampiezza* dell'onda e il numero $kx - \omega t + \phi$ si chiama la *fase*. Le due onde

$$A \sin(kx - \omega t + \phi)\,, \qquad A \sin(kx - \omega t + \psi)$$

(con gli stessi valori di ω e di k) si dicono *sfasate* di $\phi - \psi$ e questo numero si chiama *sfasamento* tra le due onde[5].

Per semplicità di scrittura, poniamo ora $\phi = 0$. L'onda che si trova all'istante $t = 0$ è l'onda

$$f(x) = \sin kx\,,$$

periodica di periodo $2\pi/k$. Però, i fisici riservano i termini "periodo" e "frequenza" alla variabile tempo. Il numero

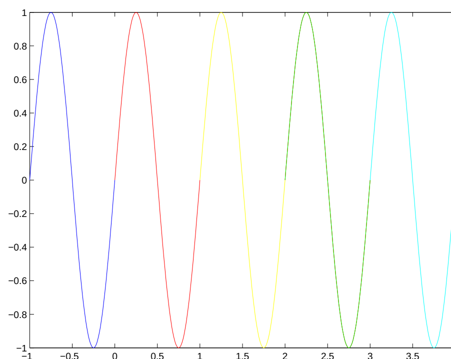
$$\lambda = \frac{2\pi}{k}$$

si chiama invece *lunghezza d'onda*. Infatti, l'onda prende i medesimi valori nei punti che distano l'uno dall'altro di $2\pi/k$. Questo può visualizzarsi in questo modo: consideriamo la funzione

$$f_e(x) = \sin kx\,, \qquad x \in [0, 2\pi/k) = [0, \lambda)\,.$$

chiamiamola[6] "onda elementare". Allora il grafico di $f(x)$ si può pensare ottenuto giustapponendo i grafici di $f_e(x + n(2\pi/k))$ per ogni numero intero n. La figura 4.15 illustra ciò, usando colori diversi per vedere i grafici che si sono giustapposti.

Figura 4.15: Lunghezza d'onda



Il numero di onde elementari che si ripetono nell'unità di spazio è 1 diviso per la lunghezza dell'onda elementare, che è $k/2\pi$. Il valore

$$\frac{1}{\lambda} = \frac{k}{2\pi}$$

[5] più comunemente, per sfasamento si prende un numero dell'intervallo $[0, 2\pi)$. Ciò può farsi perché le funzioni $\cos x$ e $\sin x$ sono periodiche di periodo 2π.

[6] Questo non è un termine usuale!

si chiama *numero d'onde* mentre k si chiama *numero d'onde angolare*.

Come si è detto, i termini "periodo", "frequenza" e "freqenza angolare" si riservano alla dipendenza dal tempo t dell'onda. Ossia, si riservano alla funzione

$$t \to A \sin(kx - \omega t + \phi)$$

con x fissato. La terminologia è quella già illustrata nel contesto del moto armonico.

Indicando con T il periodo, si ha

$$T = \frac{2\pi}{\omega}, \qquad f = \frac{1}{T} = \frac{\omega}{2\pi}$$

Il numero ω, che è la velocità angolare del moto armonico, nello studio delle onde armoniche si chiama *frequenza angolare*.

Studiamo infine la velocità di fase, scrivendo

$$\sin(\omega t + kx) = \sin k \left(\frac{\omega}{k} t + x \right) .$$

La velocità di fase, ossia la velocità di spostamento dell'onda, è

$$v = \frac{\omega}{k} = \frac{\lambda}{T} .$$

Onde stazionarie

Le onde si "propagano" nello spazio, al trascorrere del tempo; questa almeno sembra l'informazione che può ricavarsi dalle considerazioni precedenti. Però, in uno stesso mezzo possono propagarsi onde diverse. Consideriamo due onde armoniche con la medesima ampiezza

$$f(x - vt) = A \cos k(x - vt) , \qquad g(x - vt) = A \cos m(x - \nu t) .$$

Se le due onde hanno il medesimo numero d'onde, $k = m$ e la medesima velocità, $v = \nu$, allora nel mezzo si vede un'onda di ampiezza doppia,

$$2A \cos(kx - vt) .$$

Altrimenti, si trova la funzione

$$(x, t) \longrightarrow A \left[\cos k(x - vt) + \cos m(x - \nu t) \right] .$$

Usando le formule di prostaferesi si trova la funzione

$$\begin{aligned} & 2A \cos \left(\frac{(k+m)x - (kv + m\nu)t}{2} \right) \cos \left(\frac{(k-m)x + (m\nu - kv)t}{2} \right) \\ & = 2A \cos \left(\frac{k+m}{2} \left[x - \frac{kv + m\nu}{k+m} t \right] \right) \cos \left(\frac{k-m}{2} \left[x + \frac{m\nu - kv}{k-m} t \right] \right) \end{aligned}$$

Supponiamo ora che la velocità di propagazione sia uguale per le due onde, $v = \nu$. In questo caso si trova il segnale

$$\left[2A \cos \frac{k-m}{2} (x - vt) \right] \cos \frac{k+m}{2} (x - vt)$$

Interpretiamo questa formula: quando i due segnali hanno numero d'onde molto simile, $k \sim m$ e "molto grande", allora $\epsilon = k - m$ è un numero piccolo relativamente sia a k che ad m. Tenendo fissa la posizione x, al variare del tempo in questa posizione si vede passare l'onda $\cos \frac{k+m}{2}(x - vt) \sim \cos k(x - vt)$, modulata dal segnale di bassa frequenza $2A \cos \frac{k-m}{2}(x - vt)$.

Se l'onda in questione è un'onda sonora, nella posizione x si sente un suono di frequenza circa uguale a $kv/2\pi$, ed ampiezza $2A \cos[(k-m)(x-vt)/2]$ lentamente variabile nel tempo, con legge periodica di "piccola" frequenza $(k - m)v/(4\pi)$.

Si ricordi che la minima frequenza dei suoni udibili dall'orecchio umano è 16 Herz. Se accade che

$$\frac{(k-m)v}{4\pi} > 16$$

l'orecchio sente comparire un "terzo suono", diverso dai due suoni emessi dagli strumenti, ed in generale sgradevole. Questo suono si chiama appunto *terzo suono* o *suono di Tartini*.

Il caso estremo si ha quando le due onde hanno il medesimo numero d'onde e la medesima velocità, ma si propagano in verso opposto, $v = -\nu$. In tal caso si trova

$$2A \cos kx \cos kvt\,. \tag{4.15}$$

In questo caso niente si propaga: la somma di onde uguali propagantesi in versi opposti genera una configurazione $A \cos kx$, stazionaria, "modulata" dal segnale sovrapposto $\cos kvt$.

Nonostante che in questo caso niente si propaghi, per conservare la memoria dell'origine di questo fenomeno, si chiama ancora "onda" la funzione (4.15) e, per sottolineare che "niente si propaga", la chiamiamo *onda stazionaria*.

Interferenza

Consideriamo ora due onde uguali, ossia con la medesima ampiezza, numero d'onde e velocità di propagazione, ma sfasate l'una rispetto all'alta. Ossia consideriamo le due onde

$$f(x - vt) = A \cos k(x - vt)\,, \qquad g(x - vt) = A \cos k(x - vt + \phi)\,.$$

Se queste due onde si propagano nel medesimo mezzo, il segnale propagantesi nel mezzo è

$$A\left[\cos k(x - vt) + \cos k(x - vt + \phi)\right] = \left[2A \cos \frac{k\phi}{2}\right] \cos\left(kx - vt + \frac{\phi}{2}\right)\,.$$

Si ha quindi un segnale di ampiezza $2A \cos k\phi/2$. che può essere nulla.

In pratica, la differenza di fase è dovuta alla differente lunghezza del cammino percorso dalle due onde; e quindi in realtà ϕ varia cambiando posizione, $\phi = \phi(x)$. Se l'onda in questione è un'onda sonora, in un punto x fissato si sente un suono uguale a quello descritto da $f(x)$, ma con una diversa intensità. L'intensità varia al variare della posizione d'ascolto e in certe posizioni può annullarsi. Questo fenomeno si chiama *interferenza* delle due onde.

4.7.3 Appendice: Funzioni omogenee

Una funzione $f(\mathbf{r})$, con $\mathbf{r} \in \mathbf{R}^n$, si chiama *positivamente* *omogenea* *di grado* ν (anche non intero) se

$$f(t\mathbf{r}) = |t|^\nu f(\mathbf{r})\,.$$

Supponiamo che una funzione omogenea di grado ν sia di classe C^1 per $\mathbf{r} \neq 0$. Derivando rispetto a t i due membri e calcolando le derivate per $t = 1$, si trova l'uguaglianza

$$\nabla f(\mathbf{r}) \cdot \mathbf{r} = \nu f(\mathbf{r}) \qquad \forall \mathbf{r}\,. \tag{4.16}$$

Questa relazione si chiama *identità di* *Eulero* e caratterizza le funzioni omogenee e di classe C^1 (per $\mathbf{r} \neq 0$) di grado ν. Infatti, consideriamo il caso $t > 0$. Derivando la funzione

$$t^{-\nu} f(t\mathbf{r})$$

si trova

$$\frac{\mathrm{d}}{\mathrm{d}t} t^{-\nu} f(t\mathbf{r}) = -\nu t^{-\nu-1} f(t\mathbf{r}) + t^{-\nu}(\nabla f(t\mathbf{r}) \cdot \mathbf{r})$$
$$= t^{-\nu-1}\left[-\nu f(t\mathbf{r}) + \nabla f(t\mathbf{r}) \cdot (t\mathbf{r})\right]\,.$$

Per ipotesi la (4.16) vale per ogni $\mathbf{r}$; in particolare vale nel punto $t\mathbf{r}$, ossia vale

$$\nu f(t, \mathbf{r}) = \nabla f(t, \mathbf{r}) \cdot (t\mathbf{r})\,;$$

e quindi

$$\frac{\mathrm{d}}{\mathrm{d}t} t^{-\nu} f(t\mathbf{r}) = 0\,.$$

Dunque,

$$t^{-\nu} f(t\mathbf{r}) = c\,, \qquad \text{ossia} \qquad f(t\mathbf{r}) = ct^{\nu}$$

ove c è una opportuna costante. Ponendo $t = 1$ si trova $c = f(\mathbf{r})$ e quindi

$$f(t\mathbf{r}) = t^{\nu} f(\mathbf{r})\,.$$

In modo analogo si procede per $t < 0$ e quindi si trova che la funzione è omogenea di grado ν.

4.7.4 Appendice: La dimostrazione del teorema 132

Dim. Per semplicità di notazioni, proviamo l'asserto con $x_0 = 0$, $y_0 = 0$. Fissiamo i punti x ed y e consideriamo la funzione di una sola variabile

$$\phi(t) = f(tx, ty)\,.$$

Si noti che $\phi(1) = f(x, y)$ mentre $\phi(0) = f(0, 0)$.

Scriviamo

$$\phi(t) = \{f(tx, ty) - f(tx, 0)\} + \{f(tx, 0) - f(0, 0)\}\,.$$

Consideriamo la funzione della sola variabile t

$$m(t) = f(tx, 0)\,.$$

In quest'espressione, il numero x si considera come un parametro il cui valore è fissato. Applichiamo ad essa la seconda formula degli incrementi finiti: esiste $c \in (0, 1)$ tale che

$$f(x, 0) - f(0, 0) = m(1) - m(0) = \frac{\mathrm{d}}{\mathrm{d}t} m(t)_{|t=c} = x f_x(cx, 0)\,.$$

Si noti che si sono usate in questo calcolo solamente proprietà delle funzioni di una variabile e che, per applicare la seconda formula degli incrementi finiti, si è usata l'esistenza della derivata di $f(tx,0)$ che, a sua volta, implica la continuità di $f(tx,0)$.

Il numero c dipende dalla particolare funzione $m(t)$ usata e quindi, in definitiva, dipende da x. Dato che $c \in (0,1)$, si ha $|cx| < |x|$. Usando ciò, stimiamo

$$\frac{|x\,[f_x(cx,0)-f_x(0,0)]\,|}{\sqrt{x^2+y^2}} \le |f_x(cx,0)-f_x(0,0)|\ .$$

La condizione $|cx| < |x|$ e la continuità della derivata prima mostrano che il limite per (x^2+y^2) tendente a zero è nullo. Dunque vale

$$f(x,0)-f(0,0) = xf_x(0,0)+\mathrm{o}(x,y)\,. \tag{4.17}$$

Consideriamo ora la funzione $n(\tau)$, dipendente dai parametri x ed y,

$$n(\tau) = f(x,\tau y)\,.$$

Applicando a questa la seconda formula degli incrementi finiti, si trova $d \in (0,1)$ tale che

$$f(x,\tau y)-f(x,0) = n(1)-n(0) = \frac{\mathrm{d}}{\mathrm{d}\tau}n(\tau)_{|_{\tau=\tilde{c}}} = yf_y(x,dy)\,.$$

Si noti che il numero d dipende dalla particolare funzione con cui si lavora, e quindi dai parametri x ed y che la identificano. E' però un numero compreso tra 0 ed 1.

La precedente uguaglianza vale per ogni $\tau \in (0,1)$, in particolare essa vale se come valore di τ si sceglie 1:

$$f(x,y)-f(x,0) = yf_y(x,dy) = yf_y(0,0)+y\,[f_y(x,dy)-f_y(0,0)]\ .$$

L'ultima differenza verifica

$$\frac{|y\,[f_y(x,dy)-f_y(0,0)]|}{\sqrt{x^2+y^2}} \le |f_y(x,dy)-f_y(0,0)|$$

e quindi tende a zero per $\mathrm{dist}((x,y),(0,0))$ tendente a zero, per la continuità di f_y e perchè $0<d<1$. Dunque si può scrivere

$$f(x,y)-f(x,0) = yf_y(0,0)+\mathrm{o}(x,y)\,. \tag{4.18}$$

Sommando la (4.17) e la (4.18) si trova la formula cercata. ∎

Osservazione 157 Si noti che per la dimostrazione è sufficiente sapere che la funzione $f(x,y)$ è derivabile ed ha derivate continue in un intorno di (x_0,y_0). ∎

4.7.5 Appendice: dimostrazione del teorema di Schwarz

La dimostrazione del teorema di Schwarz è importante anche perché fornisce un'ulteriore rappresentazione delle derivate miste. Limitiamoci a provare il teorema assumendo $(x_0,y_0)=(0,0)$.

Mostriamo che, se $f_{x,y}(x,y)$ ed $f_{y,x}(x,y)$ sono continue in un intorno di $(0,0)$ allora si ha

$$f_{xy}(0,0) = \lim_{(h,k)\to(0,0)} \frac{f(h,k) - f(h,0) - f(0,k) + f(0,0)}{hk} = f_{y,x}(0,0)\,. \qquad (4.19)$$

L'uguaglianza implica in particolare che le derivate miste sono uguali.

Introduciamo la funzione $\phi(x) = f(x,k) - f(x,0)$. Vale:

$$\phi(h) - \phi(0) = f(h,k) - f(h,0) - f(0,k) + f(0,0)\,.$$

Dal *Teorema di Lagrange* si ha

$$\phi(h) - \phi(0) = h\phi'(c) = h[f_x(c,k) - f_x(c,0)]\,.$$

Il numero c dipende sia da h che da k, $c = c(h,k)$ e verifica

$$0 < c(h,k) < k\,.$$

Si usi ora la continuità della derivata $f_x(x,y)$, l'esistenza in ogni punto della derivata $f_{xy}(x,y)$ e di nuovo il *Teorema di Lagrange* per scrivere

$$\phi(h) - \phi(0) = h[f_x(c,k) - f_x(c,0)] = hk[f_{xy}(c(h,k),d)]\,.$$

Il punto d dipende da k e da c, quindi in definitiva da h e da k, $d = d(h,k)$, e verifica

$$0 < d(h,k) < k\,.$$

Dunque, dal *Teorema di confronto* per i limiti,

$$\lim_{(h,k)\to(0,0)} c(h,k) = 0\,, \qquad \lim_{(h,k)\to(0,0)} d(h,k) = 0$$

così che, per la continuità della derivata seconda mista

$$\lim_{(h,k)\to(0,0)} \frac{\phi(h) - \phi(0)}{hk} = \lim_{(h,k)\to(0,0)} f_{xy}(c(h,k), d(h,k)) = f_{xy}(0,0)\,.$$

Ciò prova la prima uguaglianza in (4.19). Per provare la seconda uguaglianza si introduca la funzione

$$\psi(y) = f(h,y) - f(0,y)\,.$$

Si procede in modo analogo a quanto fatto per la funzione $\phi(x)$ e si trova:

- $\psi(k) - \psi(0) = f(h,k) - f(0,k) - f(h,0) + f(0,0)$
  $= f(h,k) - f(h,0) - f(0,k) + f(0,0) = \phi(h) - \phi(0)$;
- $\lim_{(h,k)\to(0,0)} \frac{\psi(k)-\psi(0)}{hk} = f_{yx}(0,0) = \lim_{(h,k)\to(0,0)} \frac{\phi(h)-\phi(0)}{hk} = f_{x,y}(0,0)$

e quindi $f_{xy}(0,0) = f_{yx}(0,0)$. ∎

Osservazione 158 Si noto che:

- per la dimostrazione del teorema di Schwartz basta che la funzione sia di classe $C^1(V)$, con V intorno di (x_0,y_0); che esistono le derivate seconde miste in ogni punto di V e che queste sono continue in (x_0,y_0). Le derivate $f_{xx}(x,y)$, $f_{yy}(x,y)$ non intervengono.
- Supponiamo che la funzione dipenda da tre o più variabili e che x ed y denotino due delle variabili. Nel fare le derivate rispetto ad x ed y, tutte le altre variabili vengono tenute costanti; e quindi il teorema di Schwarz vale per funzioni di classe C^1, qualunque sia il numero di variabili da cui la funzione dipende; e per tutte le coppie (x,y) di variabili tali che le rispettive derivate seconde miste siano continue.

Capitolo 5

Funzioni implicite ed estremi vincolati

I termini "curva" o "superficie" hanno vari significati, tra loro interdipendenti. E' comodo conoscere da subito il significato di curva o superficie parametrica: diremo che una curva è *definita parametricamente* quando è espressa mediante un'equazione

$$x = x(t)\,, \quad y = y(t)\,, \qquad t \in (a, b)$$

(se la curva è in $\mathbf{R}^3$ c'è anche una terza componente, $z = z(t)$). Una superficie è definita parametricamente quando è definita mediante una trasformazione da una regione Ω di $\mathbf{R}^2$ in $\mathbf{R}^3$,

$$x = x(u, v)\,, \quad y = y(u, v)\,, \quad z = z(u, v)\,, \qquad (u, v) \in \Omega\,.$$

Osserviamo che ogni grafico di funzione è una curva, o una superficie, definita parametricamente. Infatti, considerando la funzione $f(x)$ della variable $x \in (a, b)$, il suo grafico è identificato dalle equazioni

$$x = t\,, \quad y = f(t) \qquad t \in (a, b)\,.$$

In modo analogo, se la funzione dipende da due variabili x ed y, il suo grafico è la superficie parametrica

$$x = u\,, \quad y = v\,, \quad z = f(u, v)\,.$$

Se accade che una curva è grafico di una funzione $y = y(x)$ oppure $x = x(y)$, diciamo che γ è una *curva cartesiana*. In modo analogo si definiscono le *superfici cartesiane* come quelle superfici che sono grafici di funzioni di due variabili.

Chiameremo "curva" anche l'insieme immagine della parametrizzazione. Al Cap. 6 saremo più precisi su questo punto.

D'altra parte, si sa che la geometria analitica definisce curve e superfici mediante equazioni: l'equazione $x^2 + y^2 = R^2$ definisce una circonferenza di raggio R (se $R > 0$; altrimenti definisce un solo punto). In questo paragrafo vogliamo dare condizioni perché un'equazione definisca una "curva" o una "superficie" in un senso che spegheremo, e vogliamo studiare problemi di massimo e di minimo "vincolati" a tali curve o superfici.

5.1 Insiemi di livello

Sia $F(\mathbf{r})$ una funzione definita su una regione $\Omega \subseteq \mathbf{R}^n$ ed a valori in $\mathbf{R}$ e si voglia studiare l'equazione $F(\mathbf{r}) = c$.

Gli insiemi

$$\mathcal{F}_c = \{\mathbf{r} \mid F(\mathbf{r}) = c\}$$

si chiamano *insiemi di livello* della funzione $F(\mathbf{r})$. Più precisamente, l'insieme $\mathcal{F}_c$ si chiama l'insieme di livello c.

Notiamo esplicitamente che l'insieme di livello è un sottinsieme del dominio della funzione e non del suo grafico. Per esempio se $\mathbf{r} = (x, y)$ l'insieme di livello si ottiene concettualmente con i tre passi seguenti:

- si costruisce il grafico della funzione, che è in $\mathbf{R}^3$;
- si taglia il grafico col piano $z = c$;
- si proietta ortogonalmente la sezione ottenuta sul piano (x, y), ottenendo l'insieme $\mathcal{F}_c$.

L'insieme $\mathcal{F}_c$ può avere la natura più varia, come mostrano gli esempi seguenti, nei quali $\Omega = \mathbf{R}^2$ e $c = 0$:

- **Esempio 1.** Se $F(x, y) = x^2 + y^2 + 1$, l'insieme $\mathcal{F}_0$ è vuoto.
- **Esempio 2.** Se $F(x, y) = x^2 + y^2$, l'insieme $\mathcal{F}_0$ è costituito dal solo punto 0.
- **Esempio 3.** Se la funzione $F(x, y)$ è identicamente nulla, $\mathcal{F}_0$ è $\mathbf{R}^2$;
- **Esempio 4.** Se $F(x, y) = (\text{sgn} x) + 1$ allora $\mathcal{F}_0$ è il semipiano $\{(x, y) \mid x < 0\}$;
- **Esempio 5.** Se $F(x, y) = 1 + (\text{sgn} y)(\text{sgn}\, x)$ l'insieme di livello è l'unione del secondo e quarto quadrante (assi coordinati esclusi).
- **Esempio 6.** Se $F(x, y) = y - x^2$ allora $\mathcal{F}_0$ è la parabola $y = x^2$, e quindi è una curva definita parametricamente da $x = t$, $y = t^2$.
- **Esempio 7.** se $F(x, y) = x^2 + y^2 - 1$ allora $\mathcal{F}_0$ è la circonferenza $x^2 + y^2 = 1$. Quest'insieme è anche immagine della curva parametrizzata da

$$x = \cos t\,, \quad y = \sin t\,, \qquad t \in [0, 2\pi)\,.$$

Esaminiamo più in dettaglio l'**esempio 7.** Come si è visto, si tratta di una curva parametrica. Se (x_0, y_0) è una soluzione, ossia un punto di $\mathcal{F}_0$, allora anche $(x_0, -y_0)$ è soluzione; e quindi l'insieme delle soluzioni non è un grafico di funzione (univoca). E' però vero che se $|y_0| \neq 0$, tagliando l'insieme delle soluzioni con una striscia

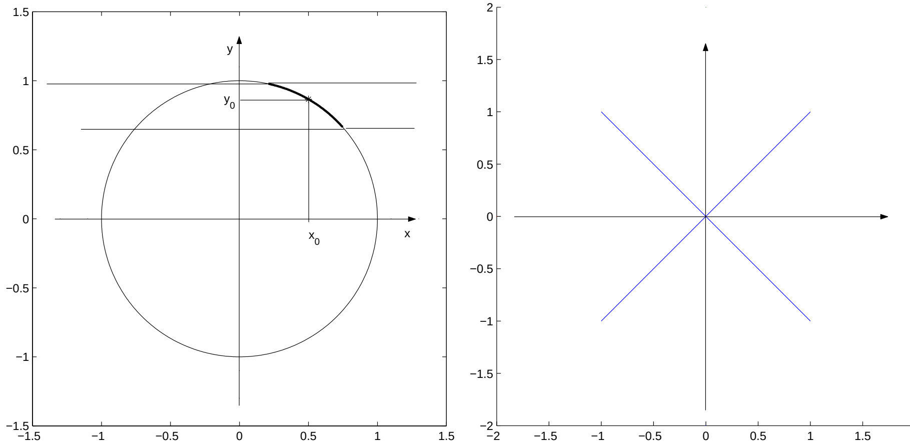
$$y_0 - \epsilon < y < y_0 + \epsilon\,,$$

con ϵ abbastanza piccolo, si trova il grafico di una funzione $y = y(x)$. Si veda la figura (5.1), a sinistra.

In questo caso particolare è facile determinare esplictamente la funzione, perché

$$y(x) = \sqrt{1 - x^2} \quad \text{se } y_0 > 0\,, \qquad y(x) = -\sqrt{1 - x^2} \quad \text{se } y_0 < 0\,.$$

Figura 5.1: Esistenza o non esistenza della funzione implicita



Se invece $y_0 = 0$, l'insieme

$$\{(x, y) \mid y_0 - \epsilon < y < y_0 + \epsilon\} \cap \{(x, y) \mid x^2 + y^2 = 1\}$$

non è grafico di una funzione $y = y(x)$. E' però grafico di una funzione $x = x(y)$. Per contrasto, vediamo l'esempio seguente:

- **Esempio 8** E' ancora $c = 0$ mentre la funzione $F(x, y)$ è

$$x^2 - y^2 = 0\,.$$

Quest'equazione è soddisfatta dai punti di ambedue le bisettrici e l'intersezione delle bisettrici con un intorno di O non è un grafico, né di una funzione $y = y(x)$ né di una funzione $x = x(y)$, si veda la figura (5.1), a destra.

Si pone quindi questo problema: supponiamo che l'insieme $\mathcal{F}_c$ sia non vuoto, e se ne conosca un suo punto $\mathbf{r_0}$. Vogliamo dare condizioni sotto le quali esiste un intorno W di $\mathbf{r_0}$ tale che $W \cap \mathcal{F}_c$ sia una curva o una superficie cartesiana. Limitandoci al caso $n = 2$ oppure $n = 3$. Se $n = 2$, vogliamo capire se l'equazione

$$F(x, y) = c$$

si può "risolvere" rispetto per esempio ad y, intendendo x come "parametro libero" ottenendo come grafico della funzione $(x, y(x))$ l'insieme $W \cap \mathcal{F}_c$; Se $n = 3$, vogliamo capire se l'equazione

$$F(x, y, z) = c$$

si può "risolvere" rispetto per esempio a z, intendendo (x, y) come "parametro libero" ottenendo come grafico della funzione $(z, z(x, y))$ l'insieme $W \cap \mathcal{F}_c$.

Quando ciò accade, si dice che l'equazione considerata *definisce implicitamente* la funzione, rispettivamente $y = y(x)$ oppure $z = z(x, y)$.

Ciò si vedrà nel prossimo paragrafo, nel quale illustreremo anche il caso di un sistema di due equazioni in tre variabili:

$$F_1(x,y,z)=c_1\,,\qquad F_2(x,y,z)=c_2\,.$$

Ci si chiede se, in un opportuno intorno di un punto (x_0,y_0,z_0), è possibile risolvere questo sistema rispetto a due "incognite" per esempio y e z, intendendo x come "parametro libero". Se ciò può farsi diremo che si è definita implicitamente una funzione, che si interpreta come curva ottenuta come intersezione di due superfici.

5.2 Il teorema della funzione implicita

Ricordiamo il problema: si ha un'equazione (o un sistema di equazioni) di cui si conosce una soluzione: si vuol sapere se l'insieme delle soluzioni è, localmente in un intorno di tale punto, una curva o una superficie cartesiana. Considereremo con qualche dettaglio il caso di equazioni $F(x,y)=c$ mentre ci limiteremo ad enunciare i risultati in due casi più generali.

5.2.1 Curve piane definite implicitamente

Ricordiamo il problema che si vuole studiare: Consideriamo l'equazione

Sia (x_0,y_0) una soluzione dell'equazione

$$F(x,y)=c\,. \tag{5.1}$$

Vogliamo dare condizioni sufficienti per l'esistenza di un intorno W di (x_0,y_0) e di una funzione $y=y(x)$ oppure $x=x(y)$ tali che

$$\{(x,y)\in W \mid f(x,y)=c\}=\{(x,y)\in W \mid y=y(x)\}$$

oppure

$$\{(x,y)\in W \mid f(x,y)=c\}=\{(x,y)\in W \mid x=x(y)\}\,.$$

La condizione che stiamo cercando è data dal teorema seguente:

Teorema 159 (della funzione implicita) Teorema della funzione implicita *Sia $F(x,y)$ una funzione di classe $C^1(\Omega)$ e sia $(x_0,y_0)\in\Omega$. Se*

$$\nabla F(x_0,y_0)\neq \mathbf{0}$$

esiste un intorno W di (x_0,y_0) tale che

$$W\cap\{(x,y)\mid F(x,y)=F(x_0,y_0)\}$$

è grafico di una funzione $y=y(x)$, oppure $x=x(y)$. Più precisamente, se $F_y(x_0,y_0)\neq 0$ allora l'equazione (5.1) definisce implicitamente una funzione $y=y(x)$ di classe C_1, e vale

$$y'(x_0)=-\frac{F_x(x_0,y_0)}{Ff_y(x_0,y_0)}\,;$$

Se $F_x(x_0, y_0) \neq 0$ allora l'equazione (5.1) definisce implicitamente una funzione $x = x(y)$ di classe C_1, e vale

$$x'(y_0) = -\frac{F_y(x_0, y_0)}{F_x(x_0, y_0)}.$$

Se ambedue le componenti di $\nabla F(x_0, y_0)$ sono non nulle, allora l'equazione $F(x, y) = f(x_0, y_0)$ definisce implicitamente sia una funzione $y = y(x)$ che una funzione $x = x(y)$.

Presentiamo (una parte della) dimostrazione di questo teorema, fissando l'attenzione sul caso $F_y(x_0, y_0) \neq 0$. Il punto (x_0, y_0) è uno dei punti nel quale vale l'uguaglianza (5.1), ossia si ha

$$F(x, y) = c = F(x_0, y_0). \tag{5.2}$$

In tal caso, proviamo

Si ha:

Teorema 160 *Valga $F(x_0, y_0) = c$. Sia $F(x, y)$ di classe C^1 e sia $F_y(x_0, y_0) \neq 0$. Sotto queste condizioni,* esistono *un intorno U di x_0 ed un intorno V di y_0 ed* esiste *un'*unica *funzione $y = y(x)$ definita in U ed a valori in V che ha per grafico l'insieme (5.2) ossia tale che*

- $y(x_0) = y_0$,
- $F(x, y(x)) = c$ *per ogni* $x \in U$,
- $y(x) \in V$.

Questa funzione è di classe C^1 e inoltre

$$y'(x_0) = -\frac{F_x(x_0, y_0)}{F_y(x_0, y_0)}. \tag{5.3}$$

Dim. Per seguire questa dimostrazione, si guardi la figura 5.2. In questa figura, $x_0 = y_0 = 3$ e il punto (x_0, y_0) è indicato con un asterisco. Per il *teorema della permanenza del segno,* vale $F_y(x, y) > 0$ in un intorno W di (x_0, y_0). Siano $\tilde{U}_0$ un intorno di x_0 e $\tilde{V}_0$ un intorno di y_0 tali che se $x \in \tilde{U}_0$, $y \in \tilde{V}_0$ allora $(x, y) \in W$.

Limitiamoci a considerare le coppie (x, y) con $x \in \tilde{U}_0$, $y \in \tilde{V}_0$. Siano y_1, y_2 elementi di $\tilde{V}_0$ tali che

$$y_1 < y_0 < y_2.$$

Consideriamo la funzione $\phi(y) = F(x_0, y)$. Questa funzione è continua e strettamente crescente e $\phi(y_0) = c$. Dunque, $\phi(y_1) < c$, $\phi(y_2) > c$.

Il *Teorema della permanenza del segno* mostra l'esistenza di un intorno $U \subseteq \tilde{U}_0$ di x_0 tale che

$$F(x, y_1) < c, \qquad F(x, y_2) > c \qquad \forall x \in U;$$

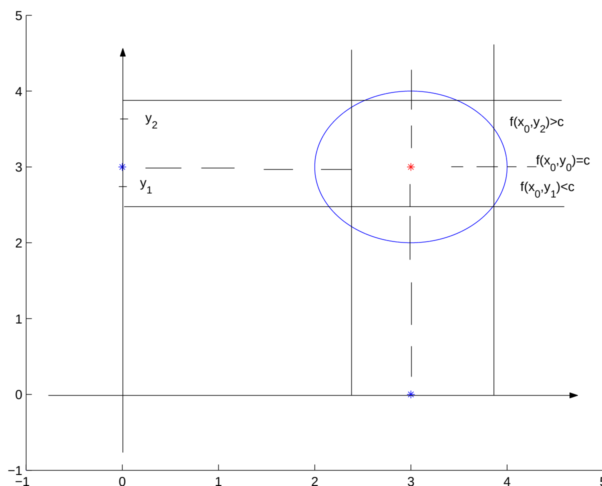
e, per ogni fissato $x \in U$, la funzione $y \to F(x, y)$ è strettamente crescente. Dunque, per x fissato, esiste un unico numero $y = y(x) \in \tilde{V}$ tale che

$$F(x, y(x)) = c.$$

Ciò prova l'esistenza della funzione $y(x)$.

Omettiamo la dimostrazione della regolarità.

Figura 5.2: La dimostrazione del teorema della funzione implicita



Accettando il fatto non provato che $y(x)$ è derivabile, si derivino i due membri dell'uguaglianza

$$F(x, y(x)) = 0\,.$$

Si trova, per il Teorema 135,

$$0 = F_x(x, y(x)) + F_y(x, y(x))y'(x)$$

e quindi

$$y'(x) = -\frac{F_x(x, y(x))}{F_y(x, y(x))}\,.$$

Calcolando per $x = x_0$ si trova la (5.3) ∎

La funzione $y = y(x)$ così costruita si dece *definita implicitamente* dall'equazione (5.1).

5.2.2 Superfici definite implicitamente

Consideriamo ora l'equazione

$$F(x, y, z) = c = F(x_0, y_0, z_0)\,. \tag{5.4}$$

Vogliamo dare condizioni sotto le quali sia possibile considerare due delle tre variabili come "parametri liberi" e risolvere rispetto alla terza, ottenendo quindi per esempio una *funzione implicita*

$$z = z(x, y)$$

che interpretiamo come equazione cartesiana di una superficie.

Limitiamoci ad enunciare il teorema che dà una condizione solamente sufficiente per l'esistenza della funzione implicita.

Teorema 161 *Sia $F(\mathbf{r}, z)$ una funzione a valori reali della variabile $(\mathbf{r}, z) \in \mathbf{R}^n$ (z indica l'ultima componente del vettore $(\mathbf{r}, z)$ e pertanto è un numero reale).*

Supponiamo che la funzione sia di classe C^1 e che valga

$$F(\mathbf{r_0}, z_0) = c\,, \qquad F_y(\mathbf{r_0}, z_0) \neq 0\,.$$

Esiste un intorno W di $(\mathbf{r_0}, z_0)$ ed esiste un'unica funzione $z = z(\mathbf{r})$ tale che

$$\{(\mathbf{r}, z) \in W \mid F(\mathbf{r}, z) = c = F(\mathbf{r_0}, z_0)\} = \{(\mathbf{r}, z) \in W \mid z = z(\mathbf{r})\}\,.$$

La funzione $z(\mathbf{r})$ è di classe C^1 ed il suo gradiente è (indicando con x^1,...,x^{n-1} le componenti di $\mathbf{r}$)

$$-\frac{1}{F_z(\mathbf{r_0}, z_0)} \begin{bmatrix} F_{x^1}(\mathbf{r_0}, z_0) & F_{x^2}(\mathbf{r_0}, z_0) & \dots & F_{x^{n-1}}(\mathbf{r_0}, z_0) \end{bmatrix}'\,. \quad \blacksquare$$

Nel caso particolare $\mathbf{r} = (x, y)$ la funzione che si trova è

$$z = z(x, y)\,,$$

ossia la rappresentazione parametrica di una superficie cartesiana.

Osservazione 162 Il teorema precedente si può applicare ad una qualsiasi delle variabili, purché la derivata parziale relativa sia non nulla; per esempio, si potrà applicare alla prima invece che all'ultima componente. ∎

5.2.3 Curve intersezione di due superfici

Studiamo ora un sistema di due equazioni in tre incognite. Consideriamo prima di tutto un esempio:

Esempio 163 Studiamo il problema

$$x^2 + y = 1\,, \qquad y - x + z = 0\,.$$

Questo sistema può scriversi come

$$z = x^2 + x - 1\,, \qquad y = 1 - x^2\,,$$

ossia il sistema ha infinite soluzioni, una per ogni valore di x. Diremo che questo sistema definisce implicitamente le due funzioni $z(x) = x^2 + x - 1$ e $y(x) = 1 - x^2$ o meglio diremo che definisce una funzione della variabile reale x, a valori in $\mathbf{R}^2$. ∎

Consideriamo in generale il sistema delle due equazioni in tre incognite

$$f(x, y, z) = c\,, \qquad g(x, y, z) = d\,. \tag{5.5}$$

Supponiamo che (x_0, y_0, z_0) risolva questo sistema e, fissato x_0, consideriamo la funzione $\mathbf{F}(y, z)$ da $\mathbf{R}^2$ in $\mathbf{R}^2$

$$\mathbf{F}(y, z) = \begin{bmatrix} F(x_0, y, z) \\ G(x_0, y, z) \end{bmatrix}\,.$$

Lo jacobiano di questa trasformazione è

$$j(x_0, y, z) = \det \begin{bmatrix} F_y(x_0, y, z) & F_z(x_0, y, z) \\ F_y(x_0, y, z) & F_z(x_0, y, z) \end{bmatrix} .$$

Vale:

Teorema 164 *Siano $F(x, y, z)$ e $F(x, y, z)$ funzioni di classe C^1 e sia*

$$j(x_0, y_0, z_0) \neq 0 .$$

Esiste un intorno W di (x_0, y_0, z_0) ed una unica funzione di classe C^1

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \phi(x) \\ \psi(x) \end{bmatrix} \tag{5.6}$$

da $\mathbf{R}$ in $\mathbf{R}^2$, tale che

$$\{(x, y, z) \in W, \text{ soluzioni di (5.5)}\} = \{(x, y, z) \mid y = \phi(x), \ z = \psi(x)\} .$$

La derivata della funzione $x \to \begin{bmatrix} \phi(x) & \psi(x) \end{bmatrix}'$ è

$$\frac{d}{dt} \begin{bmatrix} \phi(x_0) \\ \psi(x_0) \end{bmatrix} = - \begin{bmatrix} F_y(x_0, y_0, z_0) & F_z(x_0, y_0, z_0) \\ G_y(x_0, y_0, z_0) & G_z(x_0, y_0, z_0) \end{bmatrix}^{-1} \begin{bmatrix} F_x(x_0, y_0, z_0) \\ G_x(x_0, y_0, z_0) \end{bmatrix} .$$

Omettiamo la dimostrazione.

Naturalmente diremo che la funzione (5.6) è *definita* *implicitamente* dal sistema di equazioni (5.5).

5.3 Il teorema della funzione inversa ed i cambiamenti di variabili

Il *teorema della funzione inversa* si può vedere come ulteriore caso del teorema della funzione implicita, nel caso in cui l'equazione da risolvere sia

$$\mathbf{F}(\mathbf{r}) = \mathbf{x} , \qquad \mathbf{r} \in \mathbf{R}^n , \quad \mathbf{x} \in \mathbf{R}^n$$

ma conviene vederlo come studio dei *cambiamenti di variabile*.

Cominciamo ad illustrarlo nel caso più semplice $n = 1$. Abbiamo quindi una funzione $F(x)$ di una sola variabile x definita su un intervallo (a, b) ed ivi di classe C^1. Se la sua derivata non si annulla, si ha $F'(x) > 0$ oppure $F'(x) < 0$ in ogni punto di (a, b) e quindi $F(x)$ è strettamente monotona su (a, b). Dunque è invertibile.

E' naturale investigare se l'osservazione precedente possa estendersi al caso di funzioni di più variabili. Lo studio di questo problema conduce al "teorema della funzione inversa", di cui ora illustriamo l'interesse.

Abbiamo visto che talvolta conviene rappresentare i punti di $\mathbf{R}^3$ mediante coordinate sferiche oppure, a seconda delle applicazioni, cilindriche. In altri casi si usano coordinate ellittiche che, sul piano, sono date dalle trasformazioni

$$x = ra \cos\theta , \qquad y = rb \sin\theta , \qquad r > 0 , \qquad \theta \in [0, 2\pi) .$$

In generale si ha questa situazione: si hanno due regioni Ω ed Ω' di $\mathbf{R}^n$ (conviene considerarle in due "copie diverse" di $\mathbf{R}^n$) e una trasformazione invertibile $\mathbf{x} = \mathbf{F}(\mathbf{r})$ da Ω' con immagine uguale ad Ω. In questo caso i punti di Ω si possono rappresentare, invece che con le loro coordinate cartesiane, con quelle del punto $\mathbf{r}$ di Ω' che univocamente gli corrisponde.

Per esempio lavorando con coordinate polari nel piano,

$$\mathbf{r} = (r, \theta) \in (0, +\infty) \times [-\pi, \pi)\,, \qquad \mathbf{x} = (x, y)\ :\quad \begin{cases} x = r\cos\theta \\ y = r\sin\theta\,. \end{cases}$$

Questa è una trasformazione invertibile dalla striscia $\Omega = (0, +\infty) \times (-\pi, \pi)$ alla regione Ω che è il piano $\mathbf{R}^2$ privato del semiasse $\{(x, y)\,,\ x \le 0\}$.

Per molte applicazione è necessario che la trasformazione sia oltre che invertibile anche differenziabile e con inversa essa stessa differenziabile.

In pratica non è difficile riconoscere che la trasformazione con cui si lavora è differenziabile, e spesso anche riconoscere che è invertibile; è più difficile calcolare esplicitamente l'inversa e verificare che essa è differenziabile. Fortunatamente il teorema seguente dà una condizione sufficiente per l'invertibilità (si noti: solamente locale) e per la differenziabilità della funzione inversa.

Teorema 165 (teorema della funzione inversa) *Sia* $\mathbf{x} = \mathbf{F}(\mathbf{r})$ *una funzione definita su un aperto* $\Omega' \subseteq \mathbf{R}^n$ *ed a valori in* $\mathbf{R}^n$*, di classe* $C^1(\Omega')$*. Sia* $\mathbf{r_0} \in \Omega'$ *un punto in cui il determinante jacobiano* è diverso da zero:

$$\det J_{\mathbf{F}}(\mathbf{r_0}) \neq 0\,.$$

Sotto tali condizioni esistono un aperto A *contenente* $\mathbf{r_0}$ *ed un aperto* B *contenente* $\mathbf{x_0} = \mathbf{F}(\mathbf{r_0})$ *con queste proprietà:*

- *la funzione* $\mathbf{F}(\mathbf{r})$ *è biunivoca su* A*, con immagine uguale a* B*. La restrizione di* $\mathbf{F}(\mathbf{r})$ *all'aperto* A *ammette quindi funzione inversa definita sull'aperto* B*. Indichiamola col simbolo* $\mathbf{G}(\mathbf{x})$*.*
- *La funzione* $\mathbf{G}(\mathbf{x})$ *è di classe* $C^1(B)$*;*
- *vale*

$$J_{\mathbf{G}}(\mathbf{x_0}) = [J_{\mathbf{F}}(\mathbf{r_0})]^{-1}\,, \qquad \mathbf{x_0} = \mathbf{F}(\mathbf{r_0})\,. \tag{5.7}$$

Ossia, la matrice jacobiana delle funzione inversa $\mathbf{G}(\mathbf{x})$ *calcolata in* $\mathbf{x_0}$ *è* $[J_{\mathbf{F}}(\mathbf{r_0})]^{-1}$*.*

Si noti che, accettando la differenziabilità della funzione inversa, la formula per $J_{\mathbf{G}}(\mathbf{x_0})$ discende dalla formula di derivazione a catena. Infatti, sia $\mathbf{G}(\mathbf{x})$ la funzione inversa di $\mathbf{F}(\mathbf{r})$ e supponiamo di sapere che la funzione $\mathbf{G}(\mathbf{x})$ è differenziabile. Per la definizione di funzione inversa,

$$\mathbf{G}(\mathbf{F}(\mathbf{r})) = \mathbf{r}\,.$$

La matrice jacobiana della trasformazione $\mathbf{r} \to \mathbf{r}$ è I, la matrice identità. Dunque, dalla formula di derivazione a catena,

$$I = J_{\mathbf{G}}(\mathbf{x_0}) J_{\mathbf{F}}(\mathbf{r_0})\,, \qquad x_0 = \mathbf{F}(\mathbf{r_0})$$

ossia

$$J_{\mathbf{G}}(\mathbf{x_0}) = [J_{\mathbf{F}}(\mathbf{r_0})]^{-1}\,.$$

E' importante notare che il teorema della funzione inversa afferma:

1. l'invertibilità locale;
2. la regolarità dell'inversa.

Invece, l'invertibilità su tutta Ω generalmente non vale, come prova il caso delle coordinate polari nel piano. Per esse, lo jacobiano è uguale ad r e quindi non nullo per $r > 0$ e qualunque θ; ma se vogliamo una trasformazione biunivoca dobbiamo imporre a θ di appartenere ad un intervallo di lunghezza non maggiore di 2π.

Il teorema della funzione inversa ha numerose dimostrazioni, tutte interessanti. In appendice mostriamo una dimostrazione nel caso di una trasformazione da $\mathbf{R}^2$ in sé, basata sul teorema della funzione implicita.

In questo capitolo assumeremo che F sia di classe C^1 e che gli zeri di ∇F siano isolati.
Ricordiamo che i punti nei quali $\nabla F(x_0, y_0) = 0$ si chiamano *punti critici* della funzione F.

5.4 Ulteriori esempi

Ricordiamo che il *Teorema della funzione implicita* dà una condizione sufficiente perchè un insieme di livello sia localmente grafico di una funzione: se $F(x, y)$ è di classe C^1, se esiste (x_0, y_0) tale che $F(x_0, y_0) = c$ e se $F_y(x_0, y_0) \neq 0$, allora esiste un intorno W di (x_0, y_0) la cui intersezione con l'insieme di livello è grafico di una (unica) funzione

$$y = y(x)\,.$$

Analogo risultato vale se $F_x(x_0, y_0) \neq 0$. In questo caso la funzione è $x = x(y)$.

Si ricordi che questa condizione è solamente sufficiente, e non necessaria, come prova l'esempio seguente:

Esempio 166 La funzione $F(x, y) = (x - y)^2$ definisce implicitamente la funzione $x = y$. Ma in $(0, 0)$ le sue derivate sono identicamente nulle. ∎

Fissiamo un punto (x_0, y_0) tale che $\nabla F(x_0, y_0) \neq 0$. Chiamiamo *curva di livello* o *curva definita implicitamente* l'insieme

$$\gamma = \{(x, y) \mid F(x, y) = F(x_0, y_0)\}\,.$$

Si noti che potrebbero esistere punti $(x, y) \in \gamma$ nei quali il gradiente si annulla. Noi abbiamo richiesto solamente che $\nabla F(x_0, y_0) \neq 0$.

Ovviamente, ogni grafico di funzione è una curva definita implicitamente dall'equazione

$$F(x, y) = y - f(x) = 0\,.$$

Si ricordi che avevamo già notato che gni grafico di funzione è anche curva definita parametricamente.

Se $\nabla F(x_0, y_0) \neq 0$, esiste un intorno W di (x_0, y_0) tale che $W \cap \gamma$ è grafico di una funzione. Per esempio di una funzione $y = y(x)$. E' quindi possibile definire la tangente in $(x_0, y(x_0)) = (x_0, y_0)$ a tale grafico. Per definizione, la chiamiamo *tangente alla curva* γ nel punto (x_0, y_0).

La retta tangente è

$$\begin{aligned} y &= y(x_0) + y'(x_0)(x - x_0) \\ &= y(x_0) - \frac{F_x(x_0, y_0)}{F_y(x_0, y_0)}(x - x_0) \,. \end{aligned}$$

Essendo $F(x, y(x))$ identicamente nulla, la sua derivata è zero. La derivata per $x = x_0$ è

$$\nabla F(x_0, y_0) \cdot \begin{bmatrix} 1 & y'(x_0) \end{bmatrix} \,,$$

ossia, $\nabla F(x_0, y_0)$ è ortogonale alla tangente. Per definizione, diremo che il gradiente è *ortogonale alla curva*. Si ha quindi:

Dunque:

Teorema 167 *Sia γ è una curva di livello di una funzione $F(x, y)$ di classe C^1. Il gradiente di $F(x, y)$ in ciascuno dei punti di γ che non sono punti cretici di F, è ortogonale alla curva stessa.*

Per concludere, mostriamo che niente può dirsi nei punti nei quali le condizioni del teorema della funzione implicita non valgono. Abbiamo già visto (esempio 166) una funzione che non soddisfa alle condizioni del teorema della funzione implicita, ma che definisce implicitamente una funzione regolare. D'altra parte:

Esempio 168 Consideriamo l'equazione

$$F(x, y) = y^2 - x^2 = 0 \,.$$

Le due derivate parziali si annullano in $(0, 0)$.

L'equazione si risolve facilmente, trovando le soluzioni

$$y = x \,, \qquad y = -x \,.$$

La figura 5.3 a sinistra mostra che l'insieme di livello non è una curva cartesiana in nessun intorno dell'origine. A destra mostra la superficie (un paraboloide a sella) di cui si considera l'insieme di livello.

Consideriamo ora l'equazione

$$F(x, y) = y^3 - x^2 = 0$$

Ancora si annullano le due derivate parziali di $F(x, y)$ in $(0, 0)$. L'equazione definisce però un'unica funzione,

$$y = \sqrt[3]{x^2} \,.$$

Questa funzione non è derivabile in $y = 0$. Il suo grafico è nella figura 5.4, a destra mentre a sinistra è rappresentata la funzione di cui si calcola la curva di livello. ∎

Figura 5.3: Il grafico della funzione $y = \sqrt[3]{x^2}$

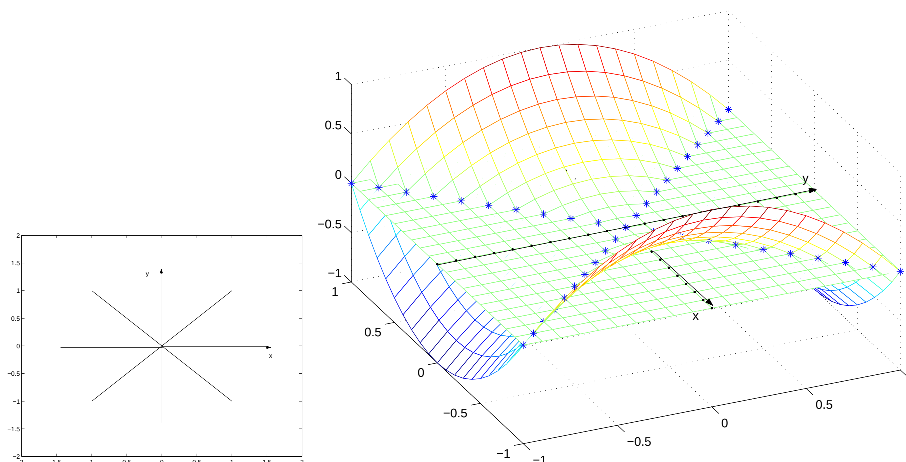


Figura 5.4: Esiste la funzione implicita anche se le condizioni del teorema non sono soddisfatte

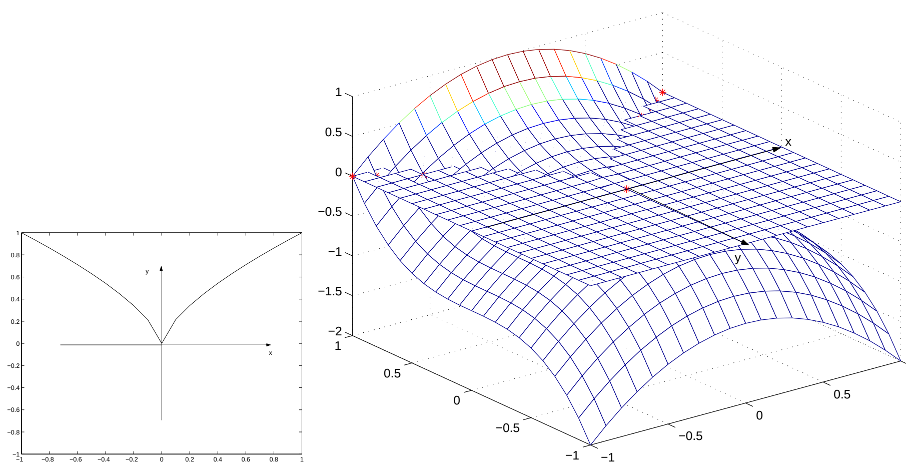
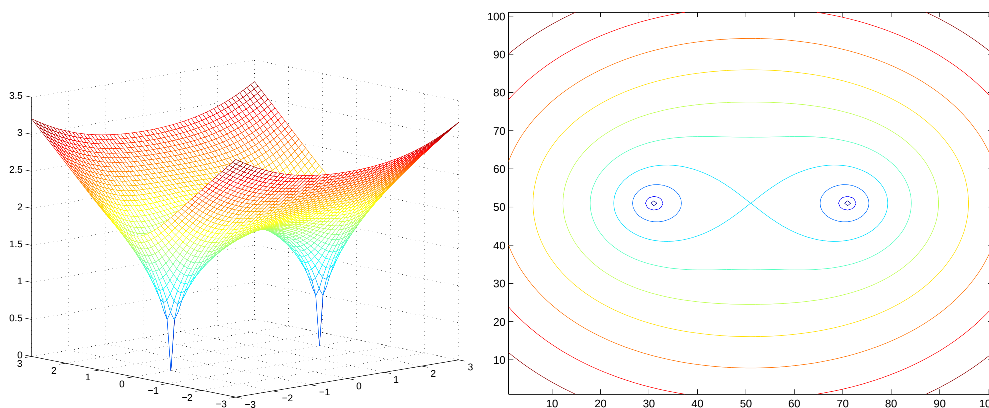


Figura 5.5: Il grafico della funzione (5.8)



Infine, consideriamo l'esempio seguente

Esempio 169 La funzione che si considera è

$$f(x,y) = \{[(x-2)^2 + y^2][(x+2)^2 + y^2]\}^{1/5}\,, \tag{5.8}$$

il cui grafico è riportato nella figura 5.5, a sinistra. A destra si riportano alcune delle sue curve di livello, corrispondenti a varie quote.

Si noti che, per quote basse, la "curva di livello" si spezza in due curve e per quota uguale a 0 si riduce a due punti (i due punti di minimo della funzione).

La funzione ha un punto di sella di coordinate (0,0, 1). La curva di livello 1 ha un "punto doppio" in $(0,0)$. ∎

Il teorema della funzione implicita asserisce che se $\nabla f(x_0, y_0) \neq 0$ (ed $f(x,y)$ è di classe C^1) allora in un intorno di (x_0, y_0) la curva di livello

$$f(x,y) = f(x_0, y_0)$$

è grafico di una funzione $y = y(x)$ oppure $x = x(y)$. Si sa già che curva di livello può essere un grafico anche se $\nabla f(x_0, y_0) = 0$. L'esempio seguente ribadisce questo fatto.

Esempio 170 La funzione è

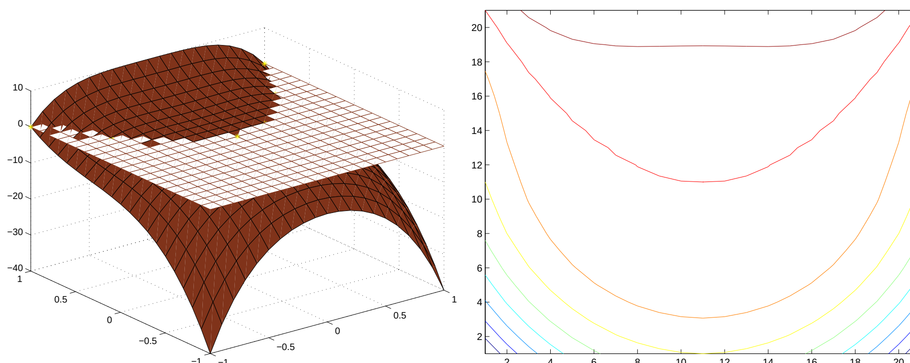
$$f(x,y) = y^3 - x^2y^2 + x^2y - x^4$$

e il punto (x_0, y_0) è $(0,0)$. E': $\nabla f(0,0) = 0$. Ciò nonostante, la curva di livello $f(x,y) = f(0,0) = 0$ è la parabola $y = x^2$ perché

$$f(x,y) = (y - x^2)(y^2 + x^2)\,.$$

La figura 5.6 mostra a sinistra il grafico di $z = f(x,y)$ e la sua intersezione col piano $z = 0$; a destra varie curve di livello della funzione.

Figura 5.6: Gradiente nullo ma curva di livello regolare



5.4.1 Superfici assegnate in modo implicito e curve intersezione di due superfici

Così come le curve, anche le superfici possono assegnarsi in modo implicito. Sia $F(\mathbf{r}) = F(x, y, z)$ una funzione di tre variabili, di classe C^1, e si voglia risolvere l'equazione

$$F(x, y, z) = c\,.$$

Sia $\mathbf{r}_0 = (x_0, y_0, z_0)$ un punto che verifica l'equazione. Dal *Teorema delle funzioni implicite* si sa che se

$$F_z(\mathbf{r}_0) \neq 0$$

allora esiste un intorno W di $\mathbf{r_0}$ ed esiste una funzione $z = \phi(x, y)$ per cui

$$\{(x, y, z) \in W \mid F(x, y, z) = c\} = \{(x, y, z) \in W \mid z = \phi(x, y)\}\,.$$

Ossia, localmente l'insieme delle soluzioni è il grafico della funzione $z = \phi(x, y)$.

Discorso analogo vale se una delle altre due derivate parziali è non nulla.

Di conseguenza, se $\nabla F(x, y, z)$ non si annulla, ogni insieme di livello di $F(\mathbf{r})$ è "fatto di tanti pezzi di grafici, e quindi di superfici". Usa considerare anche un tale insieme una *superficie* definita però in modo *implicito*.

Si abbiano ora due superfici definite in modo implicito da

$$f(x, y, z) = c\,, \qquad g(x, y, z) = d\,.$$

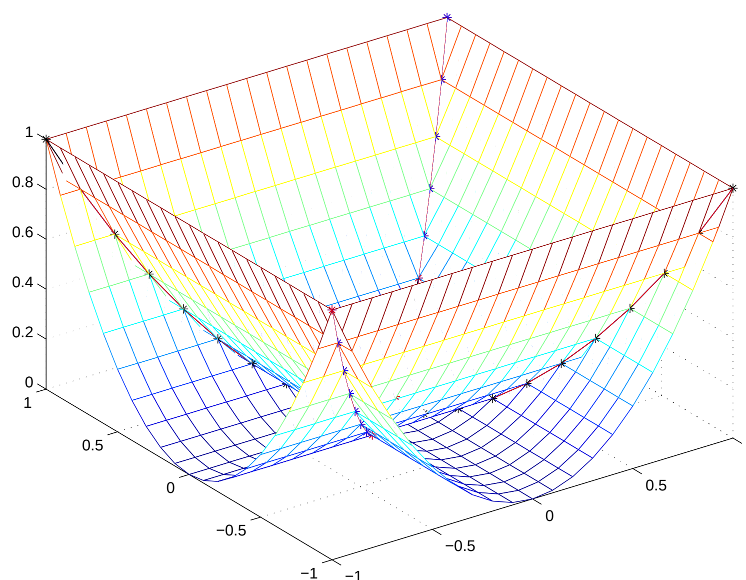
I punti (x, y, z) che appartengono all'intersezione dei due sostegni risolvono il sistema

$$f(x, y, z) = c\,, \qquad g(x, y, z) = d\,. \tag{5.9}$$

Sia (x_0, y_0, z_0) una soluzione di questo sistema. Se lo jacobiano

$$\det \begin{bmatrix} f_y(x_0, y_0, z_0) & g_y(x_0, y_0, z_0) \\ f_z(x_0, y_0, z_0) & g_z(x_0, y_0, z_0) \end{bmatrix}$$

Figura 5.7: Intersezione di due cilindri



non è nullo, si sa (dal Teorema 164) che l'insieme delle soluzioni di (5.9) che appartiene ad un opportuno intorno W di (x_0, y_0, z_0) ha forma

$$y = \phi(x)\,, \qquad z = \psi(x)$$

e quindi è una curva cartesiana di $\mathbf{R}^3$; ossia, localmente l'insieme delle soluzioni di (5.9) è una curva. Chiameremo ancora *curva* l'insieme di tali soluzioni e, più precisamente, diremo che si tratta di una curva ottenuta come intersezione di due superfici. La figura 5.7 illustra l'intersezione dei due cilindri $z = x^2$ e $z = y^2$.

5.5 Estremi vincolati

Diremo che un punto $\mathbf{r}_0$, rispettivamente o di γ o di Σ, è un *massimo relativo,* oppure *un minimo relativo,* di $g(\mathbf{r})$ *vincolato* a γ o a Σ quando è un massimo o un minimo della restrizione di $g(\mathbf{r})$ alla curva γ o, rispettivamente, alla superficie Σ.

In questo capitolo vogliamo dare condizioni necessarie per gli estremi vincolati, che estendano la condizione "derivata prima nulla nei punti estremi".

Osservazione 171 Daremo condizioni solamente necessarieche devono venir soddisfatte da un punto $\mathbf{r_0}$ di massimo o di minimo vincolato ad una curva γ o ad una superficie Σ, definite implicitamente come curve, rispettivamente superficie, di livello di una funzione di classe C^1. Quindi interessa solamente considerare il comportamento di F e di g in un intorno di $\mathbf{r_0}$. Dunque, non sarà necessario assumere che il gradiente di F sia ovunque diverso da zero.

Basterà assumere che sia $\nabla F(\mathbf{r_0}) \neq 0$. Per continuità, il gradiente $\nabla F(\mathbf{r})$ sarà diverso da zero in un intorno di $\mathbf{r_0}$. ∎

Esamineremo con qualche dettaglio il caso degli estremi vincolati ad una curva piana e, per sommi capi, il caso degli estremi vincolati ad una superficie ed ad una curva dello spazio.

5.5.1 Estremi vincolati ad una curva piana

Considerando ancora il caso di funzioni di due variabili, vogliamo studiare gli estremi di una funzione $g(x, y)$ vincolati ad una curva γ; ossia, considerando i punti di minimo[1], vogliamo studiare quei punti (x_0, y_0) del sostegno di γ con questa proprietà: esiste un intorno I di (x_0, y_0) tale che se $(x, y) \in I$ appartiene anche al sostegno di γ allora si ha

$$g(x_0, y_0) \leq g(x, y) .$$

Ovviamente, un caso è banale: se la curva γ è descritta parametricamente,

$$x = x(t) , \ y = y(t) \qquad t \in (a, b)$$

il problema si riduce a studiare i punti di minimo relativo della funzione di una sola variabile $t \to g(x(t), y(t))$. Il caso interessante è il caso in cui la curva è descritta in modo implicito.

Il problema di caratterizzare i punti di minimo ha carattere locale e, come si è detto, nell'intorno di un punto nel quale il gradiente non si annulla, ogni curva di livello può scriversi in forma parametrica. Però, in pratica, trovarne l'espressione parametrica è tutt'altro che facile. Vogliamo quindi dare una condizione necessaria soddisfatta dai punti di minimo, senza dover esplicitare la curva di livello.

Ricapitolando, è data una una funzione $F(x, y)$ di classe C^1 e un punto (x_0, y_0). Assumiamo

$$\nabla F(x_0, y_0) \neq 0 .$$

E' data una funzione $g(x, y)$ di classe C^1 e supponiamo che (x_0, y_0) sia punto di minimo della $g(x, y)$ vincolato alla curva di livello

$$F(x, y) = F(x_0, y_0) .$$

Una condizione necessaria che deve essere soddisfatta è data dal teorema seguente.

Teorema 172 *Siano $F(x, y)$ e $g(x, y)$ funzioni di classe C^1 su una regione Ω e sia $(x_0, y_0) \in \Omega$ un punto tale che*

$$\nabla F(x_0, y_0) \neq 0 .$$

Sia γ la curva di livello

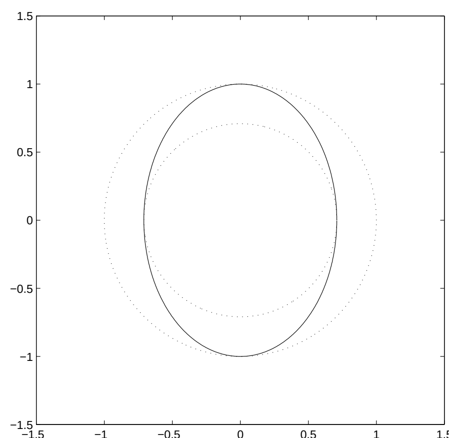
$$F(x, y) = F(x_0, y_0) .$$

Sia (x_0, y_0) un punto di massimo o di minimo di $g(x, y)$, vincolato alla curva γ. In tal caso esiste un numero λ tale che

$$\nabla g(x_0, y_0) = \lambda \nabla F(x_0, y_0) .$$

[1] i punti di massimo si trattano in modo analogo

Figura 5.8: Estremi vincolati e curve di livello



Il numero λ si chiama *moltiplicatore di Lagrange* e quando si usa il teorema precedente per la ricerca degli estremi vincolati si diche che si usa il metodo dei *moltiplicatori di Lagrange*.

Posponiamo la dimostrazione formale del teorema e presentiamone prima di tutto una giustificazione di tipo geometrico.

Considerazioni geometriche che giustificano il metodo dei moltiplicatori di Lagrange. Fissiamo l'attenzione sul punto (x_0, y_0) di γ e consideriamo la curva di livello σ di $g(x, y)$,

$$\sigma : \qquad g(x, y) = g(x_0, y_0)\,.$$

Il punto (x_0, y_0) appartiene sia a γ che a σ.

Supponiamo che la γ passi da una parte all'altra di σ. In questo caso la γ passa da una parte del piano in cui vale

$$g(x, y) < g(x_0, y_0)$$

ad una parte del piano in cui vale

$$g(x, y) > g(x_0, y_0)$$

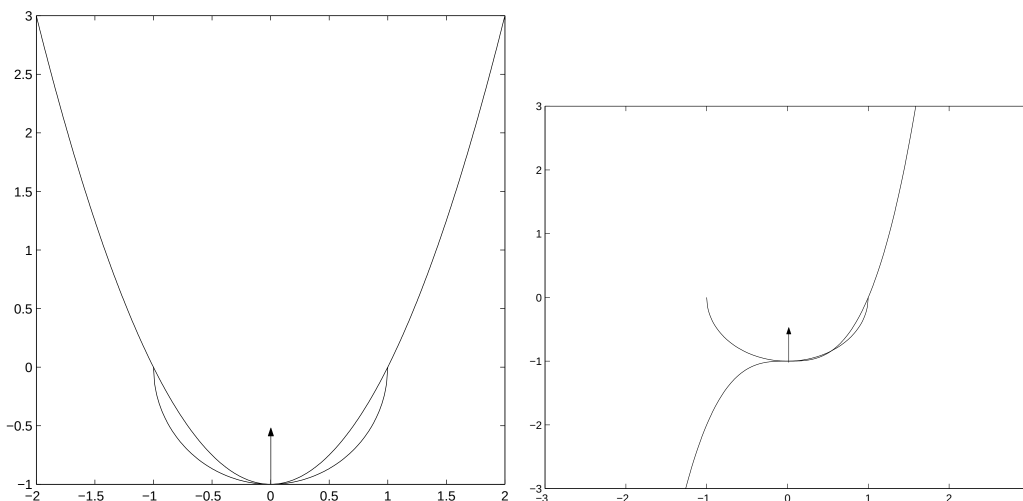
e quindi il punto (x_0, y_0) non è né di massimo né di minimo.

Dunque, nei punti di massimo e di minimo vincolati, le due curve di livello γ e σ si toccano senza attraversarsi. La fig. 5.8 illustra questo caso.

Usando gli sviluppi di Taylor si prova che le due curve γ e σ si attraversano nel loro punto comune (x_0, y_0) se le tangenti in tale punto si attraversano. Dunque nei punti di massimo e di minimo vincolato le due curve devono avere la medesima retta tangente e quindi la medesima retta normale.

Questo caso è illustrato dalla figura 5.9 a sinistra mentre la figura 5.9 a destra mostra che le due curve di livello possono attraversarsi anche nel caso in cui le tangenti coincidono.

Figura 5.9: Grafici tangenti e estremi vincolati



Sappiamo che la normale ad una curva di livello è nella direzione del gradiente della funzione e quindi nei punti di massimo oppure di minimo vincolato, $\nabla F(x_0, y_0)$ e $\nabla g(x_0, y_0)$ sono vettori colineari: esiste un numero λ tale che $\nabla g(x_0, y_0) = \lambda \nabla F(x_0, y_0)$.

Dunque, studiando il sistema della tre equazioni

$$\begin{cases} F(x,y) = 0 \\ \lambda F_x(x,y) = g_x(x,y) \\ \lambda F_y(x,y) = g_y(x,y) \end{cases} \tag{5.10}$$

nelle tre incognite λ, x ed y, si determinano dei punti tra i quali necessariamente si trovano gli estremi vincolati di $g(x, y)$.

Osservazione 173 Gli argomenti di tipo geometrico che abbiamo usato non sono rigorosi e inoltre fanno intervenire la curva di livello σ. Dunque implicitamente richiedono di lavorare in punti nei quali il gradiente della funzione $g(x, y)$ non si annulla. Il teorema vale però anche se $\nabla g(x_0, y_0) = 0$. Infatti, la dimostrazione analitica presentata più avanti non fa uso di condizioni sul gradiente di $g(x, y)$.

I punti nei quali si annulla il gradiente di $g(x, y)$ si trovano dalle (5.10) scegliendo $\lambda = 0$. ■

Alcuni esempi

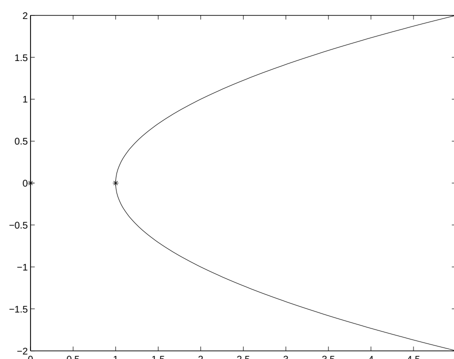
Esempio 174 Si voglia calcolare il punto su una curva piana, più vicino all'origine; ossia si voglia minimizzare sulla curva la funzione

$$g(x, y) = x^2 + y^2\,.$$

Consideriamo il caso delle quattro curve seguenti:

- la curva è implicitamente definita da
$$y^2 - x + 1 = 0$$
ed è rappresentata in figura 5.10.

Figura 5.10:



E' facile vedere geometricamente che il punto del sostegno più vicino all'origine è il punto $(1,0)$ e questo è l'unico punto per cui esiste un λ tale che
$$\begin{cases} y^2 - x + 1 = 0 \\ -1 + 2\lambda x = 0 \\ y + \lambda y = 0 \,. \end{cases}$$
Il valore di λ è $1/2$.

- Se $F(x,y) = (y^2 - x + 1)^2$ si ha il medesimo problema; ma ora il metodo dei moltiplicatori di Lagrange non è applicabile perchè le derivate parziali di $F(x,y)$ si annullano contemporaneamente. Tentando ugualmente di scrivere il sistema (5.10) si trova
$$\begin{cases} (y^2 - x + 1)^2 = 0 \\ -(y^2 - x + 1) + \lambda x = 0 \\ 2(y^2 - x + 1) + \lambda = 0 \,. \end{cases}$$
Questo sistema non dà informazioni perché scegliendo $\lambda = 0$ si vede che ogni (x,y) per cui $y^2 - x + 1 = 0$ risolve le tre equazioni.

- Sia invece $F(x,y) = x^2 + y^2 - 1$. Ovviamente, ogni punto della circonferenza minimizza la distanza. Però le due derivate parziali di $F(x,y)$ si annullano in $(0,0)$.

  Dato che $(0,0)$ non appartiene alla curva, il metodo dei moltiplicatori di Lagrange può usarsi. Il sistema (5.10) è ora
$$\begin{cases} x^2 + y^2 - 1 = 0 \\ 2x + 2\lambda x = 0 \\ 2y + 2\lambda y = 0 \,. \end{cases}$$
Ogni valore (x,y) per cui $x^2 + y^2 = 1$ risolve questo sistema (con $\lambda = -1$).

- Sia $F(x,y) = 2x^2+y^2-1$. Ancora, ambedue le derivate parziali di $F(x,y)$ si annullano in $(0,0)$, che però non appartiene alla curva. Dunque, il metodo dei moltiplicatori di Lagrange può usarsi. Il sistema (5.10) diviene:

$$\begin{cases} 2x^2 + y^2 - 1 = 0 \\ 2x + \lambda x = 0 \\ y + \lambda y = 0\,. \end{cases}$$

  Le soluzioni sono ora

$$\begin{array}{lll} x = 0 & y = \pm 1 & \text{con } \lambda = -1 \\ y = 0 & x = \pm 1/\sqrt{2} & \text{con } \lambda = -2. \end{array}$$

  Dato che $f(x,y)$ è l'ellisse in figura 5.8, i punti $(\pm\sqrt{2}, 0)$ sono di minimo mentre $(0, \pm 1)$ sono di massimo. Il metodo dei moltiplicatori di Lagrange, essendo solo basato sullo studio delle derivata prima, non permette di distinguere un caso dall'altro. ∎

Infine, applichiamo il metodo dei moltiplicatori di Lagrange per il calcolo dei punti estremi di una funzione di una sola variabile.

Esempio 175 Sia $g_0(x)$ una funzione derivabile della sola variable $x \in \mathbf{R}$. Introduciamo la funzione $g(x,y)$ di due variabili, costante rispetto ad y, data da

$$g(x,y) = g_0(x)\,.$$

Calcolare i punti estremi di $g_0(x)$ è come calcolare i punti estremi della funzione $g(x,y)$ vincolati alla curva

$$F(x,y) = 0 \qquad \text{ove } F(x,y) = y\,.$$

Si noti che la funzione $F(x,y)$ è di classe C^1 ed ha gradiente non nullo:

$$\nabla F(x,y) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Quindi i punti estremi si possono calcolare mediante il metodo dei moltiplicatori di Lagrange, ossia risolvendo il sistema seguente nelle tre incognite x, y e λ:

$$\begin{cases} F(x,y) = 0 & \text{ossia} \quad y = 0 \\ \\ \nabla g(x,y) = \lambda \nabla F(x,y) & \text{ossia} \quad \begin{cases} g_0'(x) = 0 \cdot \lambda \\ 0 = 1 \cdot \lambda \end{cases} \end{cases}$$

La prima riga impone di limitarsi a considerare punti dell'asse delle ascisse. Dall'ultima riga si vede che deve essere $\lambda = 0$. La penultima impone di guardare i valori di x per cui $g_0'(x) = 0$. Si ritrova quindi la condizione che la derivata prima deve annullarsi nei punti estremi. ∎

Dimostrazione analitica del Teorema di Lagrange

Vediamo ora una dimostrazione analitica del metodo dei moltiplicatori di Lagrange, che si presta a ulteriori generalizzazioni.

Valgano le ipotesi del teorema e sia (x_0, y_0) un punto di estremo di $g(x, y)$ vincolato alla curva di livello

$$F(x, y) = c = F(x_0, y_0)\,.$$

Dato che la funzione $F(x, y)$ è di classe C^1 e che $\nabla F(x_0, y_0) \neq 0$, l'equazione $F(x, y) = c$ definisce una curva piana, in un opportuno intorno di (x_0, y_0) che si può esprime localmente mediante l'equazione $y = y(x)$ oppure $x = x(y)$. Per fissare le idee supponiamo che valga la rappresentazione $y = y(x)$ così che

$$F(x, y(x)) = c\,. \tag{5.11}$$

Il punto $(x_0, y_0) = (x_0, y(x_0))$ è un punto di minimo vincolato per la funzione $g(x, y)$. Questo vuol dire che esiste un intorno I di x_0 tale che se $x \in I$ allora vale

$$F(x, y(x)) = 0\,, \qquad g(x_0, y_0) \leq g(x, y(x))\,.$$

Imponendo che sia nulla la derivata prima di $g(x, y(x))$ per $x = x_0$ si trova

$$0 = \frac{\mathrm{d}g(x_0, y(x_0))}{\mathrm{d}x} = g_x(x_0, y_0) + g_y(x_0, y_0)y'(x_0)\,.$$

Si ha quindi

$$0 = g_x(x_0, y_0) + g_y(x_0, y_0)y'(x_0)\,. \tag{5.12}$$

Dal Teorema della funzione implicita si sa che

$$y'(x_0) = -\frac{F_x(x_0, y_0)}{F_y(x_0, y_0)}\,.$$

Sostituendo in (5.12) si trova

$$g_x(x_0, y_0) + \frac{-g_y(x_0, y_0)}{F_y(x_0, y_0)}F_x(x_0, y_0) = 0\,.$$

Dunque, posto

$$\lambda = \frac{g_y(x_0, y_0)}{F_y(x_0, y_0)}\,,$$

si vede che

$$g_x(x_0, y_0) = \lambda F_x(x_0, y_0)\,.$$

Anche le derivate rispetto ad y verificano l'uguaglianza analoga,

$$g_y(x_0, y_0) = \lambda F_y(x_0, y_0)\,,$$

col medesimo valore di λ. Infatti,

$$g_y(x_0, y_0) - \lambda F_y(x_0, y_0) = g_y(x_0, y_0) + \frac{-g_y(x_0, y_0)}{F_y(x_0, y_0)}F_y(x_0, y_0) = 0\,.$$

Ciò completa la dimostrazione.

5.5.2 Estremi vincolati ad una superficie

Siano $F(x,y,z)$ e $g(x,y,z)$ due funzioni di classe C^1, definite in un intorno U di (x_0,y_0,z_0).

Supponiamo che sia $F(x_0,y_0,z_0)=c$.

Si ricordi che (x_0,y_0,z_0) è *punto di minimo* di $g(x,y,z)$, *vincolato* a $F(x,y,z)=c$ se

$$F(x,y,z)=c \quad \text{implica} \quad g(x_0,y_0,z_0)\le g(x,y,z)\,.$$

La definizione si estende facilmente ai punti di massimo vincolato.

Vale:

Teorema 176 *Sia (x_0,y_0,z_0) un punto di minimo o di massimo, vincolato a $F(x,y,z)=c$. Supponiamo che il gradiente di $F(x,y,z)$ non si annulli in (x_0,y_0,z_0). Allora esiste un numero λ per cui*

$$\lambda\nabla F(x_0,y_0,z_0)=\nabla g(x_0,y_0,z_0)\,. \tag{5.13}$$

Dim. Studiamo il caso del punto di minimo.

Per fissare le idee, supponiamo che sia $F_z(x_0,y_0,z_0)\ne 0$. Si espliciti $F(x,y,z)$ rispetto alla variabile z.

Il punto (x_0,y_0,z_0) è di minimo vincolato alla condizione $z-z(x,y)=0$, ossia, la funzione

$$g(x,y,z(x,y))$$

ha minimo libero in (x_0,y_0). Dunque, ambedue le sue derivate parziali sono nulle:

$$\begin{aligned}
\frac{\partial}{\partial x}g(x_0,y_0,z(x_0,y_0)) &= g_x(x_0,y_0,z(x_0,y_0))+g_z(x_0,y_0,z(x_0,y_0))z_x(x_0,y_0)=0\\
\frac{\partial}{\partial y}g(x_0,y_0,z(x_0,y_0)) &= g_y(x_0,y_0,z(x_0,y_0))+g_z(x_0,y_0,z(x_0,y_0))z_y(x_0,y_0)=0\,.
\end{aligned}$$

Avendo supposto $F_z(x_0,y_0,z_0)\ne 0$, si ha

$$z_x(x_0,y_0)=-\frac{F_x(x_0,y_0)}{F_z(x_0,y_0)}\,,\qquad z_y(x_0,y_0)=-\frac{F_y(x_0,y_0)}{F_z(x_0,y_0)}$$

ossia

$$\begin{aligned}
g_x(x_0,y_0,z_0)+\frac{-g_z(x_0,y_0,z_0)}{F_z(x_0,y_0,z_0)}F_x(x_0,y_0,z_0)=0\\
g_y(x_0,y_0,z_0)+\frac{-g_z(x_0,y_0,z_0)}{F_z(x_0,y_0,z_0)}F_y(x_0,y_0,z_0)=0\,.
\end{aligned}$$

Naturalmente vale anche

$$g_z(x_0,y_0,z_0)+\frac{-g_z(x_0,y_0,z_0)}{F_z(x_0,y_0,z_0)}F_z(x_0,y_0,z_0)=0\,.$$

Definendo

$$\lambda=-\frac{g_z(x_0,y_0,z_0)}{F_z(x_0,y_0,z_0)}$$

si trova che vale la (5.13). ∎

5.5.3 Estremi vincolati ad una curva dello spazio

Sia $F(\mathbf{r}) = F(x, y, z)$ una funzione di classe C^1 in una regione $\Omega \subseteq \mathbf{R}^3$ e sia γ una curva in Ω. Vogliamo dare una condizione necessaria che deve valere se un punto $\mathbf{r_0}$ è punto di estremo di $F(\mathbf{r})$ vincolato alla curva γ. Se la curva γ è data in forma parametrica il probema si riduce immediatamente alla ricerca degli estremi di una funzione di una sola variabile. Quindi consideriamo il caso in cui γ è data implicitamente, come intersezione di due superfici:

$$\gamma : \qquad g_1(x, y, z) = c\,, \quad g_2(x, y, z) = d\,.$$

Sia $\mathbf{r_0} = (x_0, y_0, z_0)$ un punto di γ, che è massimo oppure minimo di $F(\mathbf{r})$ vincolato a γ.

Supponiamo che in $\mathbf{r_0}$ valga la condizione del teorema della funzione implicita. Privilegiando, per esempio, la variabile x, supponiamo che si abbia

$$\det \begin{pmatrix} g_{1,y}(x_0, y_0, z_0) & g_{1,z}(x_0, y_0, z_0) \\ g_{2,y}(x_0, y_0, z_0) & g_{2,z}(x_0, y_0, z_0) \end{pmatrix} \neq 0\,. \tag{5.14}$$

Si ricordi che sotto questa condizione la curva γ si rappresenta, in un intorno di (x_0, y_0, z_0), in forma cartesiana, come

$$y = y(x)\,, \qquad z = z(x)$$

e

$$y_0 = y(x_0)\,, \qquad z_0 = z(x_0)\,.$$

Dunque, la funzione della sola variabile x

$$F(x, y(x), z(x))$$

ha un punto di estremo in x_0 e quindi la sua derivata prima è ivi nulla:

$$0 = F_x(x_0, y_0, z_0) + F_y(x_0, y_0, z_0)y'(x_0) + F_z(x_0, y_0, z_0)z'(x_0)\,.$$

Usando questa condizione, si potrebbe provare il teorema seguente:

Teorema 177 *Sia $\mathbf{r_0} = (x_0, y_0, z_0)$ punto di estremo della funzione $F(x, y, z)$ vincolato alla curva*

$$\gamma : \qquad g_1(x, y, z) = c\,, \quad g_2(x, y, z) = d\,.$$

Valga inoltre la condizione (5.14). In tal caso esistono due *numeri λ_1 e λ_2 tali che il punto $\mathbf{r_0} = (x_0, y_0, z_0)$ è punto estremale libero della funzione*

$$F(x, y, z) + \lambda_1 g_1(x, y, z) + \lambda_2 g_2(x, y, z)\,.$$

Ossia, nel punto (x_0, y_0, z_0) valgono contemporaneamente le condizioni seguenti:

$$\begin{aligned} & g_1(x_0, y_0, z_0) = c\,, \\ & g_2(x_0, y_0, z_0) = d\,, \\ & F_x(x_0, y_0, z_0) + \lambda_1 g_{1,x}(x_0, y_0, z_0) + \lambda_2 g_{2,x}(x_0, y_0, z_0) = 0\,, \\ & F_y(x_0, y_0, z_0) + \lambda_1 g_{1,y}(x_0, y_0, z_0) + \lambda_2 g_{2,y}(x_0, y_0, z_0) = 0\,, \\ & F_z(x_0, y_0, z_0) + \lambda_1 g_{1,z}(x_0, y_0, z_0) + \lambda_2 g_{2,z}(x_0, y_0, z_0) = 0\,. \end{aligned}$$

La coppia (λ_1, λ_2) si chiama ancora *moltiplicatore di Lagrange* (vettoriale) ed i due numeri si chiamano *moltiplicatori di Lagrange*.

5.5.4 Osservazione importante

Il metodo dei moltiplicatori di Lagrange dà condizioni necessarie che devono essere soddisfatte da un punto di estremo vincolato, senza necessità di esplicitare preventivamente l'equazione del vincolo. In ciascuno dei tre casi esaminati, le condizioni possono scriversi come segue: si introduce la funzione

$$L(\mathbf{r}) = F(\mathbf{r}) + \lambda \cdot g(\mathbf{r})\,.$$

Nei primi due casi esaminati, λ è un numero e $g(\mathbf{r})$ è una funzione a valori reali; nell'ultimo caso λ è un vettore a due dimensioni, $g(\mathbf{r}) = \begin{bmatrix} g_1(\mathbf{r}) & g_2(\mathbf{r}) \end{bmatrix}$ e il punto indica il prodotto scalare.

In tutti i casi la ricerca del minimo o massimo vincolato di $F(\mathbf{r})$ si riconduce alla ricerca dei punti estremali di $L(\mathbf{r})$. L'osservazione importante è questa: in generale gli estremi vincolati di $F(\mathbf{r})$ sono solamente punti estremali di $L(\mathbf{r})$. Non sono nè punti di massimo né punti di minimo di $L(\mathbf{r})$.

La funzione $L(\mathbf{r})$ sopra introdotta si chiama *lagrangiana* del problema (di minimo, oppure di massimo).

5.6 Appendice: la dimostrazione del teorema 165

La dimostrazione come conseguenza del Teorema della funzione implicita.

Come si è detto, consideriamo il caso $n = 2$.

Scriviamo in componenti la relazione

$$\mathbf{x} = (x, y) = \mathbf{F}(\mathbf{r}) = \begin{bmatrix} \phi(u,v) \\ \psi(u,v) \end{bmatrix}, \qquad \mathbf{r} = (u, v) \in \Omega'\,.$$

Ossia scriviamo quest'uguaglianza come

$$x = \phi(u, v)\,, \qquad y = \psi(u, v)\,. \tag{5.15}$$

Vogliamo considerare questa come un sistema di equazioni nelle incognite (u, v). Per ipotesi si sa che

$$\phi(u_0, v_0) = x_0\,, \qquad \psi(u_0, v_0) = y_0\,.$$

Si sa inoltre che $\phi(x, y)$ e $\psi(x, y)$ sono di classe C^1 e che

$$J_{\mathbf{F}}(\mathbf{r_0}) = \phi_u(u_0, v_0)\psi_v(u_0, v_0) - \phi_v(u_0, v_0)\psi_u(u_0, v_0) \neq 0\,. \tag{5.16}$$

Consideriamo la prima equazione in (5.15), che scriviamo come

$$0 = f(u, v, x) = \phi(u, v) - x\,. \tag{5.17}$$

La (5.16) mostra che $f_u(u_0, v_0, x_0) = \phi_u(u_0, v_0) \neq 0$ oppure $f_v(u_0, v_0, x_0) = \phi_v(u_0, v_0) \neq 0$. Sia per esempio

$$\phi_u(u_0, v_0) = f_u(u_0, v_0, x_0) \neq 0\,.$$

In tal caso si può risolvere l'equazione (5.17) rispetto alla variabile u ottenendo, in un opportuno aperto W contenente (u_0, v_0, x_0),

$$\begin{array}{l} u = U(v,x)\,, \qquad U(v_0,x_0) = u_0\,, \qquad \phi(U(v,x),v) = x\,, \\ U_v(v_0,x_0) = -\dfrac{f_v(u_0,v_0,x_0)}{f_u(u_0,v_0,x_0} = -\dfrac{\phi_v(u_0,v_0)}{\phi_u(u_0,v_0)}\,. \end{array} \tag{5.18}$$

Quando $(u, v, x) \in W$ allora si ha $(v, x) \in H$, aperto contenente (v_0, x_0) ed u appartiene ad un intorno di u_0.

Consideriamo ora l'equazione seguente, nell'aperto $H \times \mathbf{R}$:

$$0 = g(v,x,y) = \psi(U(v,x),v) - y\,.$$

Quest'uguaglianza è soddisfatta nel punto (v_0, x_0, y_0). Mostreremo in seguito che $g_v(v_0, x_0, y_0) \neq 0$. Accettando ciò, il teorema della funzione implicita mostra che l'equazione si può risolvere rispetto a v, ottenendo una funzione $V(x, y)$ di classe C^1

$$v = V(x,y)\,.$$

Questa uguaglianza vale per (v, x, y) in un opportuno aperto K contenente (v_0, x_0, y_0) e la funzione $V(x, y)$ è di classe C^1.

Sostituendo $v = V(x, y)$ nell'uguaglianza $u = U(v, x)$ (si veda la (5.18)) si ottiene la soluzione del sistema (5.15)

$$u = U(V(x,y),x)\,, \qquad v = V(x,y)$$

e, ricordiamo, le funzioni $U(x, y)$ e $V(x, y)$ sono di classe C^1.

Vediamo ora quali restrizioni sono state imposte ai punti (u, v) ed (x, y). Questi sono rappresentati dalle condizioni

$$(u,v,x) \in W\,, \qquad (v,x,y) \in H \times \mathbf{R}\,, \qquad (v,x,y) \in K\,.$$

Si identifica così un aperto di $\mathbf{R}^4$, contenente il punto (u_0, v_0, x_0, y_0). L'aperto A detto nel teorema è la proiezione di quest'aperto sul piano (u, v) mentre l'aperto B è la proiezione sul piano (x, y).

Per completare la dimostrazione, proviamo che $g_v(v_0, x_0, y_0) \neq 0$. Usando l'espressione di $U_v(v_0, x_0)$ in (5.18) si trova

$$\begin{array}{l} g_v(v_0,x_0,y_0) = \psi_u(U(v_0,x_0),v_0)U_v(v_0,x_0) + \psi_v(U(v_0,x_0),v_0) \\ \psi_u(u_0,v_0)\left[-\dfrac{\phi_v(u_0,v_0)}{\phi_u(u_0,v_0)}\right] + \psi_v(u_0,v_0) \\ = \dfrac{1}{\phi_u(u_0,v_0)}\left[\phi_u(u_0,v_0)\psi_v(u_0,v_0) - \phi_v(u_0,v_0)\psi_u(u_0,v_0)\right] \\ = \dfrac{1}{\phi_u(u_0,v_0)} J_{\mathbf{F}}(\mathbf{r_0}) \neq 0\,. \quad \blacksquare \end{array}$$

Capitolo 6

Curve e superfici

Le curve e le superfici in forma implicita sono già state incontrate. In questo capitolo studiamo le proprietà delle curve e delle superfici definite parametricamente. Saremo precisi nella definizione di curva mentre le "superfici" verranno definite in modo meno formale e preciso.

6.1 Curve parametriche

Conviene procedere per gradi nella definizione di curva. Una prima definizione, che verrà resa più precisa in seguito, è la seguente: Una trasformazione continua da un intervallo I in $\mathbf{R}^n$ si chiama *curva parametrica*.

Nella definizione di curva l'intervallo può essere chiuso o meno, limitato o meno. Se però l'intervallo è *chiuso e limitato* la curva si chiama un *arco*.

Una curva a valori in $\mathbf{R}^3$ si rappresenta in coordinate cartesiane nella forma

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \qquad t \in I\,.$$

La curva si dice *piana* quando la sua immagine appartiene ad un piano di $\mathbf{R}^3$, ossia quando esistono numeri a, b, c, d, indipendenti da t, tali che per ogni valore di t valga

$$ax(t) + by(t) + cz(t) = d\,.$$

Quando la curva è piana ed appartiene al piano $z = 0$ essa si rappresenta semplicemente come

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}\,.$$

Notazione analoga quando la curva appartiene agli altri piani coordinati.

Una curva si indica con una lettera greca minuscola:

$$\gamma : \quad t \to \mathbf{r}(t) \qquad t \in I\,.$$

In seguito noi ci limiteremo a considerare curve che hanno le seguenti proprietà di regolarità: la funzione $t \to \mathbf{r}(t)$ è derivabile su (a, b) con l'eccezione di un numero finito

di punti t_i. Si richiede che in questi punti (ed anche in a e in b se l'intervallo è limitato) esistano i limiti direzionali di $\mathbf{r}'(t)$. Inoltre si richiede che per $t \neq t_i$ si abbia $\mathbf{r}'(t) \neq 0$. Una curva con tali proprietà si chiama *regolare a tratti* e si parla di curva *regolare* quando essa è ovunque derivabile, con $\mathbf{r}'(t) \neq 0$ per ogni t.

Sia γ una curva regolare e sia $\mathbf{r}_0 = \mathbf{r}(t_0)$ un punto della sua immagine. Una almeno delle componenti di $\mathbf{r}(t)$, per esempio la prima componente $x(t)$, è invertibile in un intorno di t_0. Dunque l'immagine della restrizione di $\mathbf{r}(t)$ a tale intorno è anche immagine di una funzione di x. Si osservi che questo non implica che l'immagine di $\mathbf{r}(t)$ debba essere grafico di funzione, perché niente possiamo dire nei punti di t "lontani" da t_0. Questo è illustrato dalla figura 6.1, a sinistra, che riporta l'immagine, diciamo γ, della funzione

$$(\sin t)\mathbf{i} + t(\pi^2 - t^2)\mathbf{j}\,, \qquad t \in [-\pi, \pi]\,. \tag{6.1}$$

Si vede che quest'immagine non è grafico di funzione in nessun intorno di $(0,0)$ nonostante che la funzione $x = \sin t$ sia invertibile. La sua inversa è

$$t = \text{arc sin}\, x \tag{6.2}$$

e quindi la relazione tra x ed y è la funzione

$$y = [\text{arc sin}\, x] \left\{ \pi^2 - [\text{arc sin}\, x]^2 \right\} \,.$$

Il suo grafico è la parte spessa dell'immagine. Non esaurisce tutta la γ perché i valori di t ottenuti da (6.2) sono solamente quelli dell'intervallo $[-\pi/2, \pi/2]$.

E' appena il caso di notare che una curva, oltre che in coordinate cartesiane, può rappresentarsi, per esempio, in coordinate polari.

Esempio 178 La curva

$$x = t\cos t\,, \quad y = t\sin t\,, \qquad t > 0 \tag{6.3}$$

rappresenta una spirale, si veda la figura 6.1 a destra.

Essa può anche rappresentarsi in coordinate polari, come

$$\theta = t\,, \quad \rho = t\,, \qquad t > 0. \quad \blacksquare$$

Consideriamo una curva piana. Questa si chiama *curva cartesiana* se è rappresentata mediante una parametrizzazione della forma

$$t \to t\mathbf{i} + y(t)\mathbf{j}$$

oppure

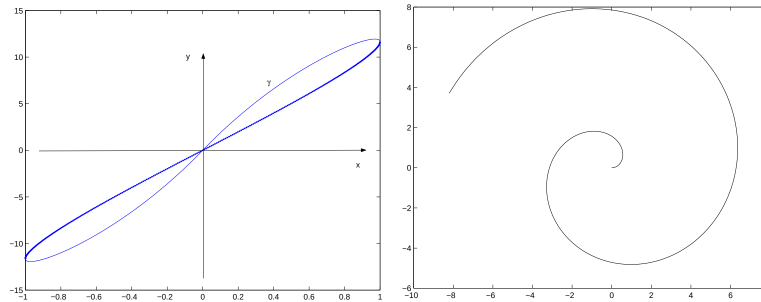
$$t \to x(t)\mathbf{i} + t\mathbf{j}\,.$$

Più in generale, se $t \to \mathbf{r}(t)$ è una curva in $\mathbf{R}^n$, si dice che questa è una curva cartesiana quando una delle componenti della funzione $\mathbf{r}(t)$ ha la rappresentazione $x_i(t) = t$.

Un arco si dice *chiuso* quando una sua parametrizzazione $\mathbf{r}(t)$, $t \in [a,b]$ verifica $\mathbf{r}(a) = \mathbf{r}(b)$.

Una curva, oppure un arco, si dice *semplice* quando $\mathbf{r}(t') = \mathbf{r}(t'')$ vale solamente per $t' = t''$ oppure se $t' = a$, $t'' = b$.

Figura 6.1: Le curve (6.1) e (6.3)



L'interpretazione fisica del concetto di curva parametrica è suggerita dai problemi della meccanica: il parametro t rappresenta il tempo ed il punto $\mathbf{r}(t)$ rappresenta la posizione all'istante t di un punto massa mobile nel tempo. Quest'interpretazione spiega la condizione di continuità posta nella definizione di curva (un punto massa non fa salti). La funzione $t \to \mathbf{r}(t)$ si chiama in fisica la *legge del moto* e il vettore $\mathbf{r}'(t)$ rappresenta la *velocità* del punto all'istante t. Si noti che la velocità può essere discontinua, per esempio quando si verificano urti.

Introduciamo ora il "verso di percorrenza" su una curva parametrica semplice: Il parametro t di una curva appartiene ad un intervallo (a, b) di $\mathbf{R}$ e $\mathbf{R}$ è un insieme ordinato. Dunque possiamo introdurre un *ordine* su una curva semplice γ di parametrizzazione $\mathbf{r}(t)$ dicendo che il punto $\mathbf{r}(t')$ viene prima del punto $\mathbf{r}(t'')$ quando $t' < t''$; ossia quando un punto mobile sulla curva traversa prima $\mathbf{r}(t')$ e poi $\mathbf{r}(t'')$. Si dice anche che, in tal caso, il punto $\mathbf{r}(t')$ *precede* $\mathbf{r}(t'')$.

Nel caso che la curva sia semplice e chiusa, il punto $\mathbf{r}(a)$ coincide col punto $\mathbf{r}(b)$ e quindi sfugge alla definizione data di ordine.

Osservazione 179 La definizione di curva parametrica è una definizione soddisfacente per alcune applicazioni della fisica, ma non per tutte, ed è del tutto insoddisfacente per la geometria. Infatti, privilegia un modo di misurare il trascorrere del tempo. Ora, due orologi diversi possono segnare ore diverse perché sono stati azzerati in istanti diversi e anche perché uno va più velocemente dell'altro. Quindi il medesimo moto viene ad avere rappresentazioni diverse, a seconda dell'orologio che si usa per descriverlo. Dobbiamo quindi migliorare la definizione di curva, tenendo conto di ciò. Osservare una proprietà cruciale del tempo: il tempo non si ferma e va in una sola direzione. Questo vuol dire che se indico con τ il tempo segnato da un orologio, al medesimo istante un secondo orologio segnerà un diverso numero, diciamo t. La corrispondenza che a τ fa corrispondere t è continua (perché il tempo non fa salti) e monotona strettamente crescente (perché il tempo non si ferma e va in una sola direzione).

Quest'osservazione è la chiave per capire la definizione generale di curva che daremo al prossimo paragrafo. ∎

6.1.1 I cambiamenti di parametro e la definizione di curva

Vogliamo ora completare la definizione di curva, tenendo conto delle ragioni esposte nell'osservazione 179.

Si dice che si cambia *parametrizzazione* della curva γ quando si opera la sostituzione $t = t(\tau)$ purché la funzione $\tau \to t(\tau)$ sia continua e strettamente crescente da un intervallo J sull'intervallo I. In pratica noi assumeremo anche che questa trasformazione sia derivabile (e quindi che sia $t'(\tau) \geq 0$). Talvolta basterà supporre che la trasformazione sia derivabile a tratti.

Rendendo più precisa la definizione di curva, si dice che $\mathbf{r}(t)$, $t \in I$ e $\tilde{\mathbf{r}}(\tau)$, $\tau \in J$ sono due diverse parametrizzazioni della stessa curva quando $\tilde{\mathbf{r}}(\tau) = \mathbf{r}(t(\tau))$ con la funzione $\tau \to t(\tau)$ strettamente crescente e suriettiva.[1]

Questa definizione corrisponde al concetto fisicamente intuitivo che una stessa "curva" può descriversi con leggi orarie diverse.

Sia ora

$$\gamma : \quad t \to \mathbf{r}(t) , \qquad t \in (a,b)$$

e si consideri la trasformazione

$$t \to b + a - t , \qquad t \in (a,b) .$$

Questa trasformazione è monotona decrescente e quindi la curva di parametrizzazione

$$t \to \mathbf{r}(b - a - t) , \qquad t \in (a,b)$$

è diversa dalla γ. Intuitivamente, la seconda curva si ottiene "percorrendo la γ all'indietro". Quando si effettua questa trasformazione sul parametro della curva, si dice che "si è cambiato il verso di percorrenza della curva" e la curva così ottenuta a partire dalla γ si indica col simbolo

$$-\gamma . \tag{6.4}$$

Ora, alcune proprietà che dovremo studiare cambieranno al cambiare della parametrizzazione, ossia della legge del moto. Altre non dipenderanno dalla parametrizzazione. Le considereremo proprietà "geometriche" della curva. Vediamo alcuni casi:

Teorema 180 *Parametrizzazioni diverse della medesima curva hanno la stessa immagine.*

L'immagine comune a tutte le parametrizzazioni di una curva γ si chiama il *sostegno* di γ. Dunque il sostegno è una proprietà geometrica della curva. Lo stesso dicasi della proprietà di essere curva chiusa o curva semplice:

[1]in modo più rigoroso: si introduce una relazione di equivalenza tra due curve definite parametricamente $\mathbf{r}(t)$ $t \in I$ e $\tilde{\mathbf{r}}(\tau)$, $\tau \in J$: esse sono equivalenti quando esiste una trasformazione continua e strettamente crescente $t(\tau)$ da J su I tale che $\mathbf{r}(t(\tau)) = \tilde{\mathbf{r}}(\tau)$ per ogni $\tau \in (\alpha, \beta)$. E quindi si definisce *curva* una classe di equivalenza rispetto a tale relazione. Per provare che quella introdotta è effettivamente una relazione di equivalenza va ricordato che la funzione inversa di una funzione crescente è essa stessa crescente.

Teorema 181 *Siano I e J due intervalli e siano $\mathbf{r}(t)$, $t \in I$, $\tilde{\mathbf{r}}(\tau)$ due curve parametriche. Sia $t(\tau)$ una trasformazione strettamente monotona da J in I, suriettiva, tale che*

$$\tilde{\mathbf{r}}(\tau) = \mathbf{r}(t(\tau))\,.$$

Allora, la curva parametrica $\mathbf{r}(t)$ è chiusa se e solo se la curva parametrica $\tilde{\mathbf{r}}(\tau)$ lo è; $\mathbf{r}(t)$ è semplice se e solo se $\tilde{\mathbf{r}}(\tau)$ lo è.

Ossia, le proprietà di essere chiusa, o di essere semplice, non dipendono dalla particolare rappresentazione parametrica di una curva ma solo dalla curva stessa: sono quindi proprietà geometriche della curva.

E' importante notare che le proprietà appena dette non cambiano nemmeno cambiando il senso di percorrenza della curva. Ossia:

La curva γ e la curva $-\gamma$ hanno il medesimo sostegno. L'una è

- semplice
- chiusa
- semplice e chiusa

se e solo se l'altra lo è.

Inoltre,

Teorema 182 *l'ordine sulla curva non muta cambiando parametrizzazione.*

E' proprio per ottenere ciò che si è imposto che i cambiamenti di parametro debbano essere strettamente crescenti.

L'ordine su $-\gamma$ è invece l'opposto di quello su γ.

Quando la trasformazione $t(\tau)$ da un intervallo J su un intervallo I è continua e strettamente monotona (crescente o meno) allora I è sia limitato che chiuso se e solo se J lo è. Dunque diremo che una curva è un *arco* quando una sua parametrizzazione è definita su un intervallo limitato e chiuso: la proprietà di essere un arco è una proprietà geometrica della curva e non cambia cambiando verso di percorrenza sulla curva, ossia essa è comune sia a γ che a $-\gamma$.

invece, dipendono dalla parametrizzazione sia la velocità $\mathbf{r}'(t)$ che la proprietà di essere una curva cartesiana.

Si consideri ora l'esempio seguente:

Esempio 183 le due curve

$$t \to (\cos t)\mathbf{i} + (\sin t)\mathbf{j}\,, \quad t \in (0, 2\pi] \qquad \text{e} \qquad t \to (\cos t)\mathbf{i} + (\sin t)\mathbf{j}\,, \quad t \in (0, 4\pi)$$

hanno il medesimo sostegno (la circonferenza $x^2 + y^2 = 1$).

Le due parametrizzazioni però non possono ricondursi l'una all'altra mediante un cambiamento di parametro (che deve essere strettamente crescente) perché la prima curva è semplice e l'altra non lo è. ∎

Gli archi semplici[2] hanno molte proprietà importanti. Tra queste:

Teorema 184 *Sia*

$$\gamma : \quad t \to \mathbf{r}(t)\,, \qquad t \in (a,b)$$

un arco semplice *(chiuso o meno). Esiste soltanto un* diverso *arco che ha il medesimo sostegno, e questo è l'arco $-\gamma$.*

Grazie a questo risultato, trattando di archi semplici, possiamo usare un linguaggio più informale: se si sa che un insieme S è sostegno di un'arco semplice, possiamo parlare di "arco S" intendendo uno dei due archi che hanno S per sostegno. In generale si intende anche di aver fissato un verso di percorrenza su S, e in tal caso si sceglie quello dei due archi che corrisponde a tale verso. Per esempio, sia

$$S = \{(x,y,z) \mid x^2+y^2+z^2 = 1\,, \quad x+y+z=0\}\,.$$

L'insieme S è una circonferenza nello spazio. Possiamo parlare dell'"arco S" intendendo implicitamente di considerare S come sostegno di un arco semplice e di scegliere una qualsiasi delle parametrizzazioni che corrispondono a tale arco.

Se si stabilisce un verso di percorrenza su S, si viene a scegliere uno solo dei due archi che hanno S per sostegno. Si noti però che in generale non esiste un modo unico per la scelta del verso di percorrenza e quindi questo linguaggio informale non identifica univocamente l'arco. Vedremo che questa difficoltà si risolve nel caso delle curve semplici e chiuse.

6.1.2 Lunghezza di un arco

Studiamo il problema di definire un numero che rappresenti la "lunghezza" di un arco in $\mathbf{R}^n$. Consideriamo per questo un arco (che indichiamo col simbolo γ) di parametrizzazione $t \to \mathbf{r}(t)$, $t \in [a,b]$. Si sa che la tangente al grafico della funzione $t \to \mathbf{r}(t)$ nel punto $(t_0, \mathbf{r}(t_0))$ ha equazione

$$\mathbf{r} = \mathbf{r}(t_0) + \mathbf{r}'(t_0)(t-t_0)\,.$$

Dividiamo l'intervallo $[a,b]$ mediante i punti t_n, equidistanti, $t_0 = a, \ldots, t_N = b$ e approssimiamo l'arco con tanti segmenti di tangente, si veda la figura 6.2: l'arco $\mathbf{r}(t)$, $t \in [t_i, t_{i+1}]$ si approssima mediante il segmento di tangente

$$\mathbf{r}(t_0) + \mathbf{r}'(t_0)(t-t_0)\,, \qquad t \in [t_i, t_{i+1}]\,.$$

Sommiamo le lunghezze dei singoli segmenti di tangente. Ripetendo questo procedimento per ogni N si costruisce una successione di numeri (L_N).

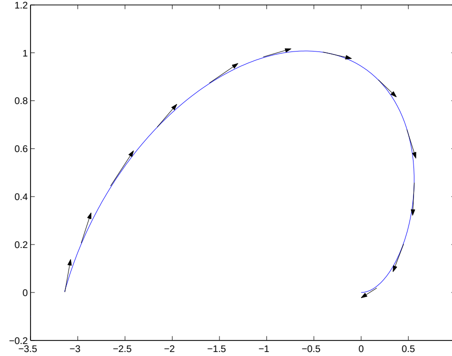
Se esiste $L = \lim L_N$, si sceglie questo numero L come "lunghezza" dell'arco γ.

Più precisamente, supponiamo che la funzione $\mathbf{r}(t)$ sia di classe C^1 e supponiamo che essa ammetta le derivate direzionali finite in ambedue gli estremi a e b. Per definire la lunghezza del'arco, si divide l'intervallo $[a,b]$ in N parti uguali mediante i punti $\frac{kT}{N}$, $0 \le k < N$ e $T = b - a$. Il segmento di tangente al grafico nel punto $(\frac{kT}{N}, \mathbf{r}(\frac{kT}{N}))$ ottenuto per $t \in (\frac{kT}{N}, \frac{(k+1)T}{N})$ ha lunghezza

$$\left\| \mathbf{r}'\left(\frac{kT}{N}\right) \right\| \cdot \frac{T}{N}\,.$$

[2] si ricordi che ogni curva chiusa è un arco.

Figura 6.2: La definizione di lunghezza: una curva e i suoi vettori approssimanti



La lunghezza totale dei segmenti di tangente è il numero

$$L_N = \sum_{k=0}^{N-1} \left|\left| \mathbf{r}' \left(\frac{kT}{N} \right) \right|\right| \cdot \frac{T}{N} \,.$$

Se esiste, il numero $L = \lim L_N$ è l'integrale della funzione $||\mathbf{r}'(t)||$ sull'intervallo $[a, b]$. Dunque, definiamo la *lunghezza* dell'arco ponendo

$$L_\gamma = \int_a^b ||\mathbf{r}'(t)|| \; \mathrm{d}t \,. \tag{6.5}$$

Osservazione 185 Al numero L siamo giunti scegliendo di dividere l'intervallo $[a, b]$ in *parti uguali.* Non è difficile mostrare che allo stesso numero L si perviene considerando una qualsiasi partizione di $[a, b]$, la cui finezza tende a zero. Ci si può chiedere però a quale numero si giunge se, invece di "approssimare" il grafico con segmenti di tangente, si sceglie di approssimarlo con segmenti di secante. E' possibile provare che si giunge al medesimo numero L, dato da (6.5). ∎

Teorema 186 *La lunghezza di un arco non muta cambiando parametrizzazione.*

Dim. Sia $t = t(\tau)$ una trasformazione crescente da $[\alpha, \beta]$ su $[a, b]$. Sia inoltre essa ovunque derivabile così che $t'(\tau) \geq 0$. Sia $\tilde{\mathbf{r}}(\tau) = \mathbf{r}(t(\tau))$.

La regola di cambiamento di variabile mostra che

$$\begin{aligned} L &= \int_a^b \left|\left| \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{r}(t) \right|\right| \mathrm{d}t = \int_\alpha^\beta \left|\left| \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{r}(t(\tau)) \right|\right| t'(\tau)\, \mathrm{d}\tau \\ &= \int_\alpha^\beta \left|\left| \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{r}(t(\tau)) t'(\tau) \right|\right| \mathrm{d}\tau = \int_\alpha^\beta \left|\left| \frac{\mathrm{d}}{\mathrm{d}\tau} \tilde{\mathbf{r}}(\tau) \right|\right| \mathrm{d}\tau \,. \quad ∎ \end{aligned}$$

Dunque, il numero che esprime la lunghezza di un arco è una proprietà geometrica dell'arco. Vale inoltre:

Teorema 187 *La lunghezza di un arco non muta cambiando il verso di percorrenza sulla curva. Ossia:*

$$L_\gamma = L_{-\gamma}\,.$$

Dim. Infatti, se $t'(\tau) < 0$ allora

$$\int_a^b \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(t)\right\|\,\mathrm{d}t = \int_\beta^\alpha \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(t(\tau))\right\| t'(\tau)\,\mathrm{d}\tau = \int_\alpha^\beta \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(t(\tau))\right\|(-t'(\tau)\,)\,\mathrm{d}\tau$$

e $-t'(\tau) = |t'(\tau)|$. Dunque anche in questo caso vale

$$L = \int_\alpha^\beta \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(t(\tau))t'(\tau)\right\|\,\mathrm{d}\tau = \int_\alpha^\beta \left\|\frac{\mathrm{d}}{\mathrm{d}\tau}\tilde{\mathbf{r}}(\tau)\right\|\,\mathrm{d}\tau\,. \quad ∎$$

Osservazione 188
- Un'interpretazione della formula (6.5) è la seguente: la parametrizzazione $t \to \mathbf{r}(t)$ si intende come legge del moto di un punto che percorre la curva. Allora, $\mathbf{r}'(t)$ è il vettore velocità del punto mobile all'istante t. La (6.5) si interpreta dicendo che l'integrale del modulo della velocità dà la lunghezza del cammino percorso.
- Sia γ una curva cartesiana, ossia

  $$\gamma:\qquad \mathbf{r}(t) = t\mathbf{i} + f(t)\mathbf{j}\,.$$

  In questo caso,

  $$\mathbf{r}'(t) = \mathbf{i} + f'(t)\mathbf{j}\,,\qquad |\mathbf{r}(t)| = \sqrt{1+[f'(t)]^2}\,. \tag{6.6}$$

  La lunghezza dell'arco ottenuto quando $t \in [a,b]$ è quindi data da

  $$L_\gamma = \int_a^b \sqrt{1+[f'(t)]^2}\,\mathrm{d}t\,. \quad ∎$$

Definiamo ora la funzione $s(t)$

$$s(t) = \int_a^t ||\mathbf{r}'(\nu)||\,\mathrm{d}\nu \qquad t \in [a,b]\,.$$

Se l'arco γ è regolare, $s'(t) > 0$ per ogni t. Ossia la trasformazione $t \to s(t)$ da $[a,b]$ su $[0,L]$ è un cambiamento di parametro per l'arco γ. Il numero $s \in [0,L]$ si chiama per questo il *parametro d'arco*. Se come parametro di γ si sceglie s si trova una nuova parametrizzazione dell'arco γ, che indichiamo con $\mathbf{r}(s)$ e che si chiama la *parametrizzazione canonica* dell'arco. La sua proprietà importante è che

$$\left\|\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{r}(s)\right\| = 1\,.$$

Infatti, essendo $s(\nu)$ la funzione inversa di $\nu(s)$, si ha:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{r}(s) &= \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{r}(\nu(s)) = \left[\frac{\mathrm{d}}{\mathrm{d}\nu}\mathbf{r}(\nu(s))\right]\nu'(s)\\ &= \left[\frac{\mathrm{d}}{\mathrm{d}\nu}\mathbf{r}(\nu(s))\right]\frac{1}{s'(\nu(s))} = \left[\frac{\mathrm{d}}{\mathrm{d}\nu}\mathbf{r}(\nu(s))\right]\frac{1}{\|(\,\mathrm{d}/\,\mathrm{d}\nu)\mathbf{r}(\nu(s))\|}\,.\end{aligned}$$

Dunque, quando la curva è data mediante la sua parametrizzazione canonica, la formula per la lunghezza si riduce a:

$$L = \int_0^L 1\,\mathrm{d}s\,.$$

Il parametro d'arco si presta allo studio delle proprietà geometriche delle curve. Però la parametrizzazione mediante il parametro d'arco è spesso piuttosto complicata e spesso non si presta a fare calcoli concreti.

Concludiamo dicendo che in fisica la funzione $s = s(t)$ si chiama *legge oraria* del moto.

6.1.3 Proprietà differenziali delle curve piane e dello spazio

Studiamo prima di tutto il caso delle curve piane.

Supponiamo che l'arco γ sia parametrizzata dalla sua lunghezza,

$$\gamma : \qquad s \to \mathbf{r}(s)\,, \qquad s \in [0, L]$$

e supponiamo che sia regolare a tratti; ossia che la funzione $s \to \mathbf{r}(s)$ sia ovunque continua; inoltre supponiamo che essa sia derivabile, con $\|\mathbf{r}'(s)\| \neq 0$ su $[0, L]$, con la possibile eccezione di un numero finito di valori $s_1,\ldots,s_k$ di s.

In tali punti richiediamo che esistano finiti[3] i limiti direzionali i limiti direzionali di $\mathbf{r}'(s)$. Il versore

$$\mathbf{t}(s) = \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{r}(s)$$

applicato nel punto $\mathbf{r}(s)$ si chiama il *versore* *tangente* alla curva nel punto $\mathbf{r}(s)$. Il versore tangente è definito salvo che in un numero finito di valori di s.

Per ogni s nel quale $\mathbf{t}(s)$ è definito, introduciamo il versore $\mathbf{n}(s)$ ortogonale a $\mathbf{t}(s)$ e diretto in modo tale che la coppia $(\mathbf{t}(s), \mathbf{n}(s))$ sia orientata positivamente (ossia, possa sovrapporsi ordinatamente ai versori $\mathbf{i}$ e $\mathbf{j}$ degli assi coordinati mediante una rotazione e una traslazione di assi). Il versore $\mathbf{n}(s)$ si chiama il versore *normale* alla curva γ.

Vale:

Teorema 189 *Sia γ una curva regolare, la cui parametrizzazione è di classe C^2. Allora il vettore $\mathbf{n}(s)$ è derivabile e in ogni punto è colineare col vettore $\mathbf{t}'(s)$.*

Dim. Infatti, $||\mathbf{t}(s)|| = 1$ per tutti gli s e quindi $\mathbf{t}(s) \cdot \mathbf{t}'(s) = 0$ per ogni s. Dunque, $\mathbf{n}(s) = \pm\frac{\mathbf{t}'(s)}{||\mathbf{t}'(s)||}$. Ciò mostra che per ogni valore di s i due vettori $\mathbf{t}(s)$ ed $\mathbf{n}(s)$ sono colineari.

Fissiamo ora un valore s_0 in cui $\mathbf{r}(s)$ ammette derivata continua e supponiamo[4] che sia $\mathbf{n}(s_0) = +\frac{\mathbf{t}'(s_0)}{||\mathbf{t}'(s_0)||}$. Ciò vuol dire che

$$\det \begin{bmatrix} \mathbf{t}(s_0) & +\frac{\mathbf{t}'(s_0)}{||\mathbf{t}'(s_0)||} \end{bmatrix} > 0\,.$$

[3]ossia, richiediamo che esistano anche i limiti direzionali di $\|\mathbf{r}'(s)\|$ e che questi siano finiti.

[4]considerazioni analoghe se $\mathbf{n}(s_0) = -\frac{\mathbf{t}'(s_0)}{||\mathbf{t}'(s_0)||}$.

Per continuità, la disuguaglianza si conserva in un intorno di s_0 e ciò mostra che in un intorno di s_0 vale

$$\mathbf{n}(s) = +\frac{\mathbf{t}'(s)}{||\mathbf{t}'(s)||}\,.$$

Il denominatore non si annulla e quindi $\mathbf{n}(s)$ è derivabile. ∎

Osservazione 190 Si ricordi che la derivata del versore $\mathbf{t}(s)$ è il limite del rapporto incrementale

$$\frac{\mathbf{t}(s+h)-\mathbf{t}(s)}{h},$$

applicato in $\mathbf{r}(s)$. ∎

Supponiamo ora di lavorare con parametrizzazioni di classe C^2 di curve regolari, e studiamo $\mathbf{t}'(s)$.

Si è visto nella dimostrazione del Teorema 189, che $\mathbf{t}'(s)$ è parallelo al versore $\mathbf{n}(s)$. Esiste quindi un numero $k(s)$ tale che

$$\mathbf{t}'(s) = k(s)\mathbf{n}(s)\,. \tag{6.7}$$

Il numero $k(s)$, che può essere positivo o negativo, si chiama la *curvatura* di γ nel punto $\mathbf{r}(s)$. Prendendo la norma dei vettori ai due membri di (6.7) si trova

$$|k(s)| = ||\mathbf{t}'(s)||\,.$$

L'esempio seguente mostra che la curvatura può cambiare segno da punto a punto di una medesima curva; e mostra anche che per calcolare tangenti, normali e curvatura non è necessario parametrizzare preventivamente la curva col parametro d'arco.

Esempio 191 Sia $\mathbf{r}(x) = (x, f(x))$ così che

$$\frac{\mathrm{d}s(x)}{\mathrm{d}x} = \sqrt{1+f'^2(x)}\,.$$

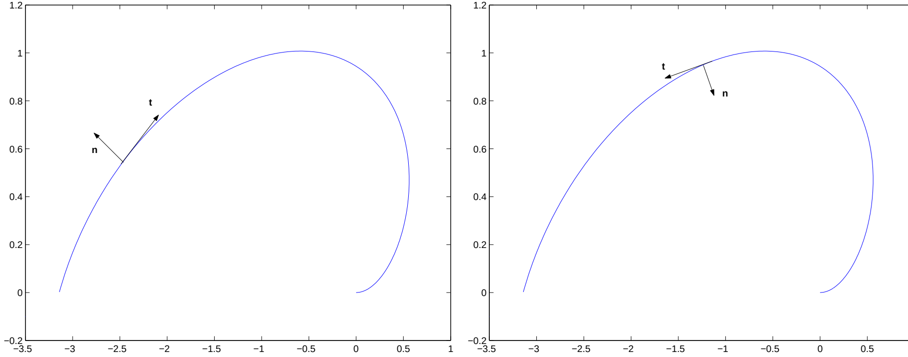
Dalla *formula per la derivata della funzione inversa*

$$\mathbf{t}(s) = \frac{\mathrm{d}\mathbf{r}(s)}{\mathrm{d}s} = \frac{\mathrm{d}r(x(s))}{\mathrm{d}x}x'(s) = \frac{1}{\sqrt{1+f'^2(s)}}\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{1}{\sqrt{1+f'^2(x)}}(1, f'(x))\right\}\,.$$

Dunque,

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{t}(x)}{\mathrm{d}s} &= \frac{1}{\sqrt{1+f'^2(x)}}\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{(1, f'(x))}{\sqrt{1+f'^2(x)}}\right] \\
& \frac{1}{\sqrt{1+f'^2(x)}}\frac{1}{1+f'^2(x)}\left[(0, f''(x))\sqrt{1+f'^2(x)} - (1, f'(x))\frac{f'(x)f''(x)}{\sqrt{1+f'^2(x)}}\right] \\
&= \frac{f''(x)}{(1+f'^2(x))^2}(-f'(x), 1) = k(x)\frac{1}{\sqrt{1+f'^2(x)}}(-f'(x), 1)\,.
\end{aligned}$$

Figura 6.3: Versore tangente e versore normale



Si trova da qui

$$k(x) = \frac{f''(x)}{(1 + f'^2(x))^{3/2}}\,.$$

La curvatura ha quindi il segno di $f''(x)$.

Si confronti con la definizione di curvatura data al par. 6.3 del testo di *Analisi Matematica 1.* ∎

Osserviamo ora che $\mathbf{t}(s) \cdot \mathbf{n}(s) = 0$ e quindi, derivando,

$$0 = \mathbf{t}'(s) \cdot \mathbf{n}(s) + \mathbf{t}(s) \cdot \mathbf{n}'(s) = k(s)\mathbf{n}(s) \cdot \mathbf{n}(s) + \mathbf{t}(s) \cdot \mathbf{n}'(s) = k(s) + \mathbf{t}(s) \cdot \mathbf{n}'(s)\,. \qquad (6.8)$$

D'altra parte, $\mathbf{n}'(s)$ è ortogonale ad $\mathbf{n}(s)$ (perché $||\mathbf{n}(s)|| = 1$) e quindi

$$\mathbf{n}'(s) = \alpha(s)\mathbf{t}(s)\,.$$

Sostituendo nella (6.8) si vede che $\alpha(s) = -k(s)$. Ne viene che le due funzioni $\mathbf{t}(s)$ ed $\mathbf{n}(s)$ risolvono il sistema di equazioni differenziali

$$\mathbf{t}'(s) = k(s)\mathbf{n}(s)\,, \qquad \mathbf{n}'(s) = -k(s)\mathbf{t}(s)\,. \qquad (6.9)$$

Queste equazioni si chiamano *Equazioni di Frenet* per le curve piane.

Il sistema di riferimento dato dai due versori $\mathbf{t}(s)$, $\mathbf{n}(s)$ (in quest'ordine) applicati nel punto $\mathbf{r}(s)$ si chiama il *riferimento mobile* sulla curva. La figura 6.3 illustra il riferimento mobile nel caso in cui la curva venga percorsa in due versi opposti.

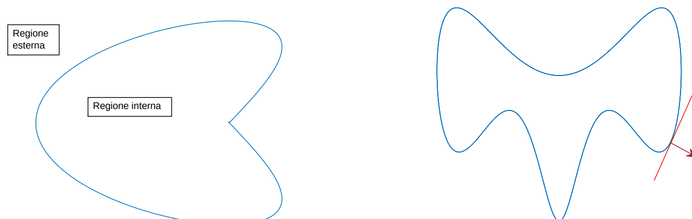
Studiamo ora il caso delle curve di $\mathbf{R}^3$. Intendiamo ancora che la curva sia parametrizzata dal parametro d'arco.

La definizione del *versore tangente* $\mathbf{t}(s)$ è ancora

$$\mathbf{t}(s) = \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{r}(s)\,.$$

Invece, la curvatura deve essere trattata in modo diverso. Assegnato il vettore $\mathbf{t}(s)$ tangente alla curva γ e di modulo 1, è ancora vero che $\mathbf{t}'(s)$ è ortogonale a $\mathbf{t}(s)$. Dunque

Figura 6.4: Le regioni interna ed esterna e la normale esterna



privilegeremo, tra le infinite direzioni normali a $\mathbf{t}(s)$, la direzione di $\mathbf{t}'(s)$; ma non c'è alcun modo di privilegiare un verso su tale direzione. Dunque decidiamo di scegliere come versore normale il versore

$$\mathbf{n}(s) = \frac{\mathbf{t}'(s)}{||\mathbf{t}'(s)||} \,. \tag{6.10}$$

Chiamiamo questo il *versore normale* alla curva γ. Chiamiamo *curvatura* il numero $k(s)$ tale che

$$\mathbf{t}'(s) = k(s)\mathbf{n}(s) \qquad \text{ossia} \qquad k(s) = ||\mathbf{t}'(s)|| = \mathbf{t}'(s) \cdot \mathbf{n}(s) \,.$$

In questo modo,

$$k(s) \geq 0 \qquad \forall s \,.$$

Osservazione 192 Se una curva regolare è piana, il suo versore normale è ovunque definito. Invece, una curva nello spazio potrebbe essere priva di versore normale su tutto un arco o addirittura ovunque. Ciò avviene se $\mathbf{t}'(s)$ è nullo. In particolare, può accadere che $\mathbf{t}'(s)$ sia identicamente zero su un intervallo. In questo caso, l'arco corrispondente è piano, parametrizzato da

$$\mathbf{r}(s) = s\mathbf{r}_0 + \mathbf{r}_1 \,.$$

La definizione di versore normale data in $\mathbf{R}^3$ non si applica in questo caso. ∎

6.2 Curve piane

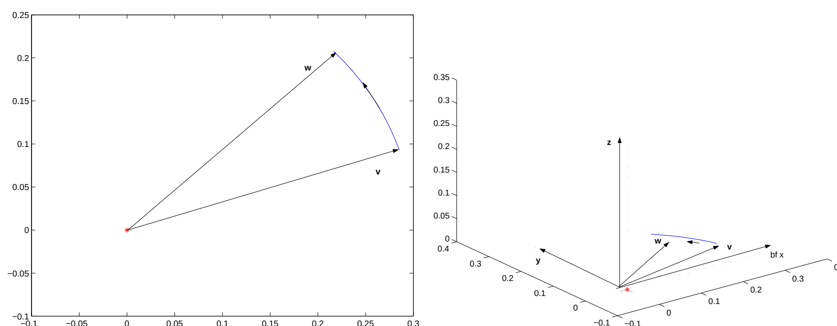
E' importante sapere che vale il teorema seguente, di enunciato del tutto intuitivo ma di dimostrazione molto complessa:

Teorema 193 (teorema di Jordan) *Sia γ una curva* piana *chiusa e semplice. Il complementare del sostegno di γ è unione di due* regioni. *Una di esse è* illimitata *(e si dice* esterna *alla curva) mentre l'altra è* limitata *e si dice la regione* interna *alla curva.*

Il teorema è illustrato nella figura 6.4, a sinistra.

Il sostegno di γ è la frontiera sia della regione interna che della regione esterna di γ (si veda il paragrafo 3.3 per la definizione di frontiera).

Figura 6.5: Regola d'Ampère per una curva piana



Esempio 194 La curva

$$x = 3\cos\theta\,, \qquad y = 2\sin\theta$$

ha per sostegno un'ellisse. La sua regione interna è la parte di piano delimitata dall'ellisse. La regione esterna è quella dei punti che "stanno fuori" dall'ellisse. ∎

Ovviamente, non esiste alcun "teorema di Jordan" per curve dello spazio!

Usa chiamare *regione di* Jordan la regione interna ad una curva piana semplice e chiusa.

Se γ indica la curva, conviene indicare con Ω_γ la sua regione interna. Vale:

Teorema 195 *Sia γ una curva semplice e chiusa. E':*

$$\Omega_\gamma = \Omega_{-\gamma}\,.$$

La regione interna ad una curva piana semplice e chiusa può essere assai complicata; ma nella maggior parte dei casi che si incontrano nelle applicazioni sarà facile identificarla.

Nel paragrafo 6.1 abbiamo usato l'ordinamento su $\mathbf{R}$ per definire un ordine sulla curva γ. Nel par. 3.2.1 abbiamo notato che il piano può venire orientato con la regola seguente: la coppia dei vettori $\mathbf{v}$ e $\mathbf{w}$ applicati in O e presi in quest'ordine, è orientata positivamente quando la semiretta identificata da $\mathbf{v}$ deve ruotare in verso *antiorario* per portarsi su quellaidentificata da $\mathbf{w}$, percorrendo l'angolo minore possibile. Si veda la figura 6.5 a sinistra.

Questa definizione può anche riformularsi mediante la *regola d'Ampère:* una persona stando in piedi nell'origine del piano xy con la testa nel verso positivo dell'asse delle quote vede la semiretta muoversi in verso antiorario, e quindi la vede passare dalla sua destra alla sua sinistra.

Sia ora γ una curva piana semplice e chiusa. Ricordiamo che il suo sostegno è sostegno, oltre che di γ, soltanto della seconda curva $-\gamma$, che si ottiene "andando all'indietro". Diciamo che γ è *orientata in modo concorde a* $\mathbf{R}^2$, o anche che è *orientata positivamente*, se vale la *regola d'Ampère:* una persona in piedi in un punto della regione interna alla curva, stando in piedi come l'asse delle quote positivo, vede un punto mobile sulla curva passare dalla sua destra alla sua sinistra. In modo equivalente, si può anche dire che un insetto che segue il punto mobile su una curva semplice e chiusa vede la regione interna alla sua sinistra, si veda la figura 6.5 a destra.

Altrimenti, diciamo che è orientata negativamente. Vale:

Teorema 196 *Delle due curve semplici e chiuse, γ e $-\gamma$, una è orientata positivamente e l'altra è oreintata negativamente.*

Supponiamo ora che la curva piana semplice e chiusa γ sia anche regolare, così che si possono definire sia il vettore tangente $\mathbf{t}(s)$ che il vettore normale $\mathbf{n}(s)$. Il vettore normale $\mathbf{n}(s)$ può puntare sia verso la regione interna che verso la regione esterna alla curva. Per il seguito avremo bisogno del vettore normale che punta verso la regione esterna alla curva γ (si veda la figura 6.4, a destra). Lo indicheremo col simbolo

$$\mathbf{n_e}(s)$$

e lo chiameremo la *normale* *esterna*.

Osservazione 197 Al paragrafo 6.1.3 si è definito il vettore $\mathbf{n}(s)$ come il vettore normale a $\mathbf{t}(s)$, orientato in modo tale che la coppia ordinata $(\mathbf{t}(s), \mathbf{n}(s))$ costituisca un sistema di riferimento positivo. Dunque $\mathbf{n}(s)$ punta verso la sinistra di $\mathbf{t}(s)$. Quando la curva è orientata positivamente, allora $\mathbf{n}(s)$ punta verso la regione interna alla curva. Quindi, in questo caso si ha $\mathbf{n}(s) = -\mathbf{n_e}(s)$. ∎

Torniamo ad usare il fatto che il sostegno di una curva semplice e chiusa è sostegno anche di una seconda curva, che si ottiene dalla prima "girando in verso opposto". Quest'osservazione permette di introdurre un linguaggio più informale, che tuttavia è limitato alle curve piane. Supponiamo che si sappia che un certo insieme del piano è il sostegno di una curva semplice e chiusa. Per esempio un quadrato o una circonferenza. Invece di scrivere esplicitamente la parametrizzazione della curva, possiamo indicare il sostegno e implicitamente intendere di scegliere quella curva semplice che ha il sostegno dato e che è orientata positivamente, senza dover esplicitamente scrivere una sua parametrizzazione. In particolare, se Ω è la regione interna ad una curva semplice e chiusa γ, e se vogliamo che γ sia orientata positivamente, potremo semplicemente indicarla come "frontiera di Ω", $\partial\Omega$.

Osservazione 198 Con questa convenzione, se γ è un sostegno di curva semplice e chiusa, si indica con γ (o, per ridondanza, $+\gamma$) la curva semplice e chiusa che ha il dato sostegno e che è orientata positivamente, e con $-\gamma$ quella che ha il dato sostegno ed è orientata negativamente. ∎

6.3 Le superfici

Studiamo ora le superfici in $\mathbf{R}^3$. Considerazioni analoghe a quello che hanno condotto a definire prima le curve parametriche e poi le curve come "oggetti geometrici" si possono ripetere per le superfici. Però sono alquanto complesse e quindi ci limiteremo a definire le superfici parametriche, mostrando quando certe proprietà che ci interessano sono indipendenti dalla parametrizzazione scelta.

6.3.1 Superfici definite parametricamente

Nel definire le curve è stato naturale partire da funzioni continue definite su intervalli. Per definire le superfici dobbiamo considerare funzioni continue di due variabili, definite quindi

su un dominio contenuto in $\mathbf{R}^2$. richiederemo che il dominio sia una regione. Una funzione continua

$$\mathbf{r}(u,v) = x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k} \tag{6.11}$$

il cui dominio è una regione di Ω si chiama una *superficie* *definita parametricamente.*

L'immagine della funzione $\mathbf{r}(u,v)$ si chiama il *sostegno* *della superficie* mentre il punto (u,v) variabile in Ω si chiama il *parametro* della superficie.

Una superficie si dice *semplice* quando valori diversi del parametro hanno per immagine punti diversi del sostegno.

Una superficie si dice *chiusa* quando il suo sostegno è la frontiera di una regione di $\mathbf{R}^3$.

Sia

$$\Sigma : \qquad (u,v) \to x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k}$$

una superficie. Diremo che la superficie è *regolare* quando la trasformazione (6.11) è di classe C^1 e inoltre la matrice jacobiana della trasformazione

$$\begin{bmatrix} x_u(u,v) & x_v(u,v) \\ y_u(u,v) & y_v(u,v) \\ z_u(u,v) & z_v(u,v) \end{bmatrix} \tag{6.12}$$

ha rango 2, ossia il massimo possibile, in ogni punto della regione Ω.

Così come nel caso delle curve, una stessa superficie può parametrizzarsi in più modi. Senza entrare in eccessivi dettagli, diremo che una trasformazione

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \to \begin{bmatrix} u(\alpha,\beta) \\ v(\alpha,\beta) \end{bmatrix}$$

è un cambiamento di parametro quando è definita su una regione $\tilde{\Omega}$, a valori in Ω; è iniettiva e suriettiva; è di classe C^1 e inoltre conserva l'orientazione di $\mathbf{R}^2$, ossia il suo jacobiano è positivo in ogni punto:

$$\det \begin{bmatrix} u_\alpha(\alpha,\beta) & u_\beta(\alpha,\beta) \\ v_\alpha(\alpha,\beta) & v_\beta(\alpha,\beta) \end{bmatrix} > 0\,.$$

Si noti l'analogia con la nozione di cambiamento di parametro per una curva. Anche nel caso delle curve il cambiamento di parametro deve conservare l'orientazione, in tal caso l'orientazione di $\mathbf{R}$.

E' ovvio che cambiando parametro non si cambia il sostegno di una superficie.

Diremo equivalenti, e le identificheremo, due superfici che differiscono solamente per la parametrizzazione.

Le funzioni definite su Ω, a valori in $\mathbf{R}^3$, sono particolari superfici, rappresentate da

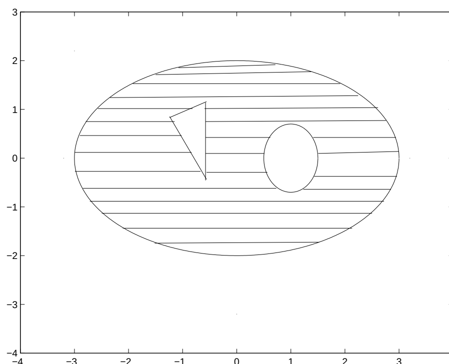
$$\Sigma : \qquad (x,y) \to x\mathbf{i} + y\mathbf{j} + z(x,y)\mathbf{k}\,.$$

Esse si chiamano *superfici cartesiane*. Il sostegno in questo caso è il grafico della funzione $z(x,y)$.

Sia ora Ω_Γ una regione di Jordan e sia $\mathbf{r}(u,v)$ una funzione continua sull'insieme chiuso costituito dall'unione della regione Ω_Γ e del supporto di Γ. In tal caso la funzione $\mathbf{r}(u,v)$ si chiama *calotta*. Ovviamente, la restrizione di $\mathbf{r}(u,v)$ ad Ω_Γ è una superficie.

Conviene estendere la definizione di calotta in questo modo. Siano date, oltre alla Γ, anche le curve di Jordan γ_1, γ_2,..., γ_n i cui sostegni non si intersecano. Supponiamo che ciascuna di queste curve abbia sostegno in Ω_Γ. Indichiamo con K l'insieme chiuso i cui punti sono quelli del sostegno di Γ e della sua regione interna Ω_Γ, esclusi i punti della regione interna a ciascuna γ_i (e quindi inclusi i punti dei sostegni delle γ_i). La figura 6.6 mostra in tratteggio un esempio di insieme K. Se la funzione $\mathbf{r}(u, v)$ in (6.11) è continua

Figura 6.6: Insieme su cui si proietta una calotta



su K essa si chiama *calotta*. Si chiama *sostegno* della calotta l'immagine della funzione $\mathbf{r}(u, v)$.

L'insieme dei punti interni di K è ancora una regione (anche se non più una regione di Jordan) e quindi la restrizione di $\mathbf{r}(u, v)$ a tale insieme è una superficie.

Parleremo di *calotta chiusa* se accade che la calotta è frontiera di un insieme aperto di $\mathbf{R}^3$.

Osservazione 199 Non si confonda il concetto di "insieme chiuso" con quello di "superficie chiusa" o di "calotta chiusa". La calotta di parametrizzazione

$$x = u\,, \qquad y = v\,, \qquad z = u + v$$

definita sul disco $u^2 + v^2 \leq 1$ ha per sostegno un insieme chiuso; ma la calotta stessa è contenuta nel piano $z = x + y$ e quindi non è una calotta chiusa. ∎

Una superficie, oppure una calotta, si indica con una lettera greca maiuscola, come per esempio Σ o Γ.

Il concetto seguente è molto delicato e noi ci limitiamo a darne una definizione grossolana. Supponiamo di avere una calotta Σ

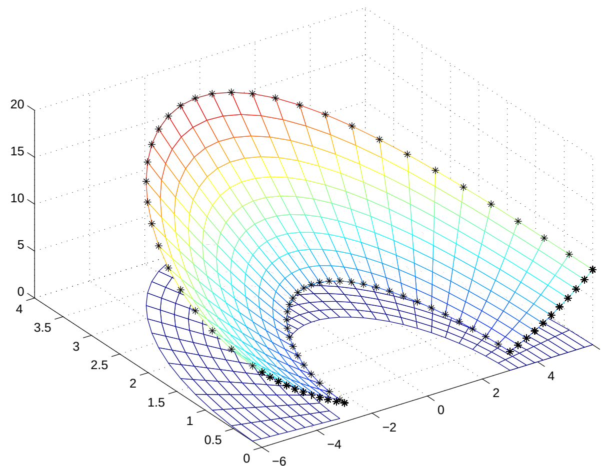
$$\mathbf{r}(u, v) = x(u, v)\mathbf{i} + y(u, v)\mathbf{j} + z(u, v)\mathbf{k}\,, \qquad (u, v) \in K\,.$$

Supponiamo che la calotta sia semplice ossia che

$$x(u', v') = x(u, v)\,, \quad y(u', v') = y(u, v)\,, \quad z(u', v') = z(u, v)$$

possa aversi solo se $u = u'$ e $v = v'$. In tal caso si chiama *bordo* della calotta Σ l'immagine della frontiera dell'insieme K; ossia l'immagine delle singole curve che delimitano l'insieme K. La figura 6.7 mostra una calotta e il suo bordo.

Figura 6.7: Una calotta ed il suo "bordo"



Esempio 200 Si è detto che la definizione di bordo è insoddisfacente. Quest'esempio ne mostra la ragione. Consideriamo la calotta definita come segue. La funzione $\mathbf{r}(\theta, v)$ è

$$\Sigma: \qquad x = \cos\theta\,, \quad y = \sin\theta\,, \quad z = v\,. \tag{6.13}$$

Il dominio della funzione è

$$0 \le v \le 1\,, \qquad \epsilon \le \theta \le 2\pi$$

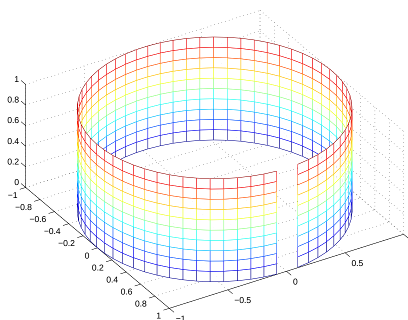
con $\epsilon > 0$. Si tratta di un cilindro a cui è stata tolta una striscetta, come in figura 6.8. Questa calotta è una calotta semplice e il suo bordo è costituito dai due archi di circonferenza (archi della circonferenza di sopra e di quella di sotto, alle quali è tolto l'arco che corrisponde a $0 < \theta < \epsilon$) e dai bordi del taglio che le congiungono.

Supponiamo ora di mandare ϵ a zero. In tal caso si trova un cilindro intero. La sua parametrizzazione non è più semplice, e quindi non possiamo più parlare di "bordo" secondo la nostra definizione, anche se appare naturale considerare le due circonferenze come il bordo del cilindro. Mentre il contributo dei due bordi del taglio "scompare".

Quest'esempio mostra che dovremmo dare un modo per definire il "bordo" anche per calotte che non sono semplici. Per esempio anche nel caso del cilindro ottenuto scegliendo $\theta \in [0, 2\pi]$. La soluzione ovvia è quella di scegliere come bordo l'immagine della frontiera dell'insieme K, in quest'esempio l'immagine del perimetro del rettangolo $[0,1] \times [0, 2\pi]$. In questo modo si otterrebbe come bordo l'insieme delle due circonferenze ed anche il segmento verticale dei punti di coordinate $(1, 0, v)$, $0 \le v \le 1$. Questa soluzione però non è accettabile. Infatti lo stesso cilindro si parametrizza anche scegliendo come dominio della funzione (6.13) l'insieme

$$0 \le v \le 1\,, \qquad -\pi \le v \le \pi$$

Figura 6.8: Ancora una calotta col suo “bordo”



e con questa parametrizzazione si trova un'altro insieme come “bordo”: l'insieme costituito dalle due circonferenze e dal segmento dei punti $(-1, 0, v)$, $0 \le v \le 1$. Ossia, il bordo così definito viene a dipendere dalla particolare parametrizzazione che si sceglie.

Ci sono vari modi per risolvere questa difficoltà: uno, più astratto, consiste nel considerare tutte le parametrizzazioni della calotta, ciascuna definita su un proprio insieme K. Si considerano quindi le immagini di tutte le frontiere di questi insiemi K e se ne fa l'intersezione. Noi seguiremo una via “più concreta” che si adatta ai casi semplici che incontreremo nelle applicazioni e che sarà illustrata al paragrafo 8.5.2. ∎

6.3.2 Il piano tangente e la normale a una superficie

Sia

$$\Sigma : \qquad (u,v) \to x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k} \qquad (u,v) \in \Omega$$

una superficie regolare e semplice. Fissiamo l'attenzione su un punto $\mathbf{r}_0$ del sostegno. Dato che la superficie è semplice, questo proviene da un unico punto (u_0, v_0) del dominio:

$$\mathbf{u}_0 = (u_0, v_0)\,, \qquad \mathbf{r}_0 = x(u_0,v_0)\mathbf{i} + y(u_0,v_0)\mathbf{j} + z(u_0,v_0)\mathbf{k}\,.$$

Consideriamo ora il segmento per $\mathbf{u}_0$, parallelo all'asse delle ascisse, ossia la curva

$$\gamma : \qquad t \to (u_0 + t, v_0)$$

Questo identifica una curva sulla superficie, parametrizzata da

$$t \to x(u_0+t, v_0)\mathbf{i} + y(u_0+t,v_0)\mathbf{j} + z(u_0+t,v_0)\mathbf{k}\,.$$

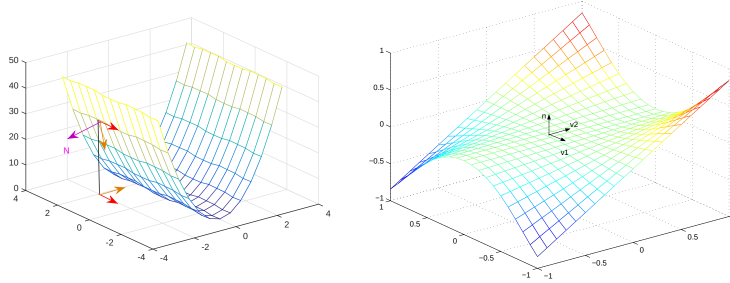
La tangente a questa curva calcolata per $t = 0$, ossia in $\mathbf{r}_0$, è identificata dal vettore $\mathbf{v_1}$ (applicato in $\mathbf{r_0}$)

$$\mathbf{v_1} = x_u(u_0,v_0)\mathbf{i} + y_u(u_0,v_0)\mathbf{j} + z_u(u_0,v_0)\mathbf{k}\,. \tag{6.14}$$

Analogamente, considerando un segmento per $\mathbf{u}_0$ parallelo all'asse delle ordinate, si trova una curva sulla superficie, la cui tangente in $\mathbf{r}_0$ è identificata dal vettore $\mathbf{v_2}$ (applicato in $\mathbf{r_0}$):

$$\mathbf{v_2} = x_v(u_0,v_0)\mathbf{i} + y_v(u_0,v_0)\mathbf{j} + z_v(u_0,v_0)\mathbf{k}\,. \tag{6.15}$$

Figura 6.9: Il piano tangente e la normale ad una superficie



In generale, il segmento

$$t \to (u_0 + at, v_0 + bt)$$

identifica una curva sulla superficie, il cui vettore tangente è identificato dal vettore

$$a\mathbf{v_1} + b\mathbf{v_2} \qquad (\text{applicato in } \mathbf{r_0})\,.$$

I vettori $\mathbf{v_1}$ e $\mathbf{v_2}$ dipendono da (u_0, v_0):

$$\mathbf{v}_1(u_0, v_0)\,, \qquad \mathbf{v}_2(u_0, v_0)$$

e sono le due colonne della matrice jacobiana (6.12) e quindi sono linearmente indipendenti (infatti abbiamo assunto che la superficie sia regolare). Dunque al variare di a e di b in $\mathbf{R}$, i vettori $a\mathbf{v_1} + b\mathbf{v_2}$ applicati in $\mathbf{r_0} = \mathbf{r}(u_0, v_0)$ descrivono un piano per $\mathbf{r_0}$, che si chiama il *piano* *tangente* alla superficie nel punto $\mathbf{r_0}$.

Si noti esplicitamente che intendiamo di scegliere come sistema di riferimento cartesiano (in generale, obliquo) in questo piano le rette identificate al vettore $\mathbf{v_1}(u_0, v_0)$ per primo e quindi $\mathbf{v_2}(u_0, v_0)$. E' su questi assi cartesiani si sceglie per verso positivo quello dei rispettivi vettori. Dunque sul piano tangente è definita un'orientazione.

Definiamo ora il *vettore normale* $\mathbf{N}(u_0, v_0)$ ponendo

$$\mathbf{N}(u_0, v_0) = \mathbf{v_1}(u_0, v_0) \wedge \mathbf{v_2}(u_0, v_0)\,.$$

Il vettore $\mathbf{N}(u_0, v_0)$ si intende applicato nel punto $\mathbf{r_0} = \mathbf{r}(u_0, v_0)$.

E' ovvio:

Teorema 201 *Se la superficie semplice Σ è regolare, sia i vettori $\mathbf{v_1}(u_0, v_0)$, $\mathbf{v_2}(u_0, v_0)$ che il vettore normale $\mathbf{N}(u_0, v_0)$ dipendono con continuità da (u_0, v_0).*

Si ricordi che la superficie si è supposta semplice. Quindi ogni $\mathbf{r_0}$ proviene da un unico punto $(u_0, v_0) \in \Omega$. Dunque in ogni punto di una superficie regolare e semplice il vettore normale $\mathbf{N}$ sopra definito è unico e questo vettore si potrà considerare come funzione del punto della superficie: $\mathbf{N} = \mathbf{N}(\mathbf{r})$. Si è così definito un campo vettoriale sulla superficie.

I concetti appena esposti sono illustrati nella figura 6.9.

Se la superficie è cartesiana si ha:

$$\mathbf{v}_1(x_0,y_0)=\begin{bmatrix}1\\0\\z_x(x_0,y_0)\end{bmatrix},\quad \mathbf{v}_2(x_0,y_0)=\begin{bmatrix}0\\1\\z_y(x_0,y_0)\end{bmatrix}\quad \mathbf{N}(x_0,y_0)=\begin{bmatrix}-z_x(x_0,y_0)\\-z_y(x_0,y_0)\\1\end{bmatrix}.$$

Per il seguito è importante ricordare la formula per $||\mathbf{N}(\mathbf{r})||$ nel caso di una superficie cartesiana:

$$||\mathbf{N}(x_0,y_0)||=\sqrt{1+[z_x(x,y)]^2+[z_y(x,y)]^2}\,. \tag{6.16}$$

Si confronti questa formula con la (6.6).

Studiamo ora come cambiano i vettori $\mathbf{v_1}(u_0,v_0)$, $\mathbf{v_2}(u_0,v_0)$ e il vettore $\mathbf{N}(u_0,v_0)$ sotto l'azione dei cambiamenti di parametro. Sia $u=u(t,s)$, $v=v(t,s)$ una trasformazione biunivoca (e di classe C^1) da una regione Ω' su Ω e consideriamo la parametrizzazione

$$\mathbf{r}(t,s)=x(u(t,s),v(t,s))\mathbf{i}+y(u(t,s),v(t,s))\mathbf{j}+z(u(t,s),v(t,s))\mathbf{k}\,.$$

Le derivate rispetto ad t ed s si calcolano mediante la regola di derivazione a catena. Poniamo:

$$a=u_t(t,s)\,,\quad b=v_t(t,s)\,,\quad c=u_s(t,s)\,,\quad d=v_s(t,s)\,.$$

Per semplicità di notazioni, scriviamo x_u invece di $x_u(u(t,s),v(t,s))$ (e analoga notazione per le derivate di y e di z, e per le derivate rispetto a v). Si ha:

$$\begin{aligned}\mathbf{r}_t(t,s)&=[x_ua+x_vb]\mathbf{i}+[y_ua+y_vb]\mathbf{j}+[z_ua+z_vb]\mathbf{k}\\ \mathbf{r}_s(t,s)&=[x_uc+x_vd]\mathbf{i}+[y_uc+y_vd]\mathbf{j}+[z_uc+z_vd]\mathbf{k}\end{aligned}$$

La componente lungo il versore $\mathbf{k}$ del prodotto vettoriale $\mathbf{r}_t(t,s)\wedge\mathbf{r}_s(t,s)$ è

$$\begin{aligned}&[(x_uy_uac+x_vy_vbd+x_uy_vad+x_vy_ubc)-(y_ux_uac+y_vx_vbd+y_ux_vad+y_vx_ubc)]\\&=(x_uy_v-x_vy_u)(ad-bc)\,.\end{aligned}$$

Proseguendo in modo analogo al calcolo delle altre componenti si trova:

Teorema 202 *Vale:*

$$\mathbf{r}_t(u(t,s),v(t,s))\wedge\mathbf{r}_s(u(t,s),v(t,s))=(ad-bc)\mathbf{r}_u(u(t,s),v(t,s))\wedge\mathbf{r}_v(u(t,s),v(t,s))\,.$$

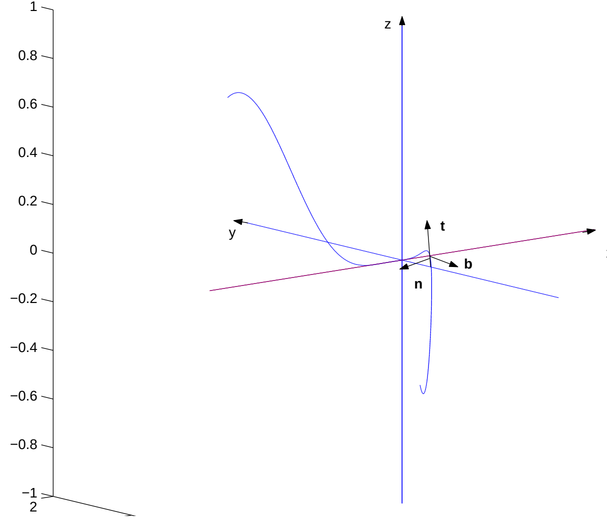
Il numero $ab-bc$ è lo jacobiano del cambiamento di parametro. Esso è positivo per la definizione di cambiamento di parametro.

Teorema 203 *effettuando un cambiamento di parametro la normale alla superficie non cambia né direzione né verso.*

Per questa ragione si dice che i cambiamenti di parametro (che hanno jacobiano positivo) lasciano invariata l' *orientazione* della superficie.

Si dice che cambiano l'orientazione della superficie quelle trasformazioni che hanno jacobiano negativo.

Figura 6.10: Riferimento mobile su una curva nello spazio



6.4 Appendici

6.4.1 Appendice: le formule di Frenet per curve nello spazio

Torniamo a considerare le curve nello spazio e completiamo le considerazioni svolte al paragrafo 6.1.3. In quel paragrafo abbiamo definito la tangente $\mathbf{t}(s)$ e la normale $\mathbf{n}(s)$ ad una curva. Il piano identificato dai vettori $\mathbf{t}(s)$ e $\mathbf{n}(s)$, applicati in $\mathbf{r}(s)$, si chiama il *piano* *osculatore* alla curva nel punto $\mathbf{r}(s)$.

Notiamo ora che $\mathbf{t}(s) \cdot \mathbf{n}(s)$ è identicamente zero e quindi ha derivata nulla. Dunque,

$$\mathbf{t}(s) \cdot \mathbf{n}'(s) = -\mathbf{t}'(s) \cdot \mathbf{n}(s) = -k(s)\,. \tag{6.17}$$

Introduciamo ora il versore $\mathbf{b}(s)$, definito da

$$\mathbf{b}(s) = \mathbf{t}(s) \wedge \mathbf{n}(s)\,.$$

Il versore $\mathbf{b}(s)$ è quindi ortogonale a $\mathbf{t}(s)$ e $\mathbf{n}(s)$ ed orientato in modo tale che la terna $(\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s))$ sia orientata positivamente. Si veda la figura 6.10 per un esempio.

Il sistema di assi cartesiano ortogonali che abbiamo descritto varia da punto a punto della curva. Per questa ragione si chiama *sistema di* *riferimento mobile* *sulla curva.*

Il vettore $\mathbf{b}(s)$ si chiama il versore *binormale* alla curva.

La derivata di $\mathbf{b}(s)$ è ortogonale a $\mathbf{b}(s)$ perché $||\mathbf{b}(s)|| = 1$ per ogni s e quindi appartiene al piano di $\mathbf{t}(s)$ e $\mathbf{n}(s)$ per ogni s. Si ha quindi

$$\mathbf{b}'(s) = \alpha(s)\mathbf{t}(s) + \beta(s)\mathbf{k}(s)\,.$$

D'altra parte,

$$\mathbf{b}'(s) = [\mathbf{t}'(s)] \wedge \mathbf{n}(s) + [\mathbf{t}(s)] \wedge \mathbf{n}'(s).$$

Ora, $\mathbf{t}'(s)$ è colineare con $\mathbf{n}(s)$ e quindi il loro prodotto vettoriale è nullo. Rimane quindi

$$\mathbf{b}'(s) = [\mathbf{t}(s)] \wedge \mathbf{n}'(s)\,.$$

Il vettore $[\mathbf{t}(s)] \wedge \mathbf{n}'(s)$ è ortogonale sia a $\mathbf{t}(s)$ che a $\mathbf{n}'(s)$. E' quindi un multiplo di $\mathbf{n}(s)$. Dunque, per ogni s esiste un numero $\tau(s)$ tale che

$$\mathbf{b}'(s) = -\tau(s)\mathbf{n}(s)\,. \tag{6.18}$$

Da qui si trova

$$\tau(s) = -\mathbf{b}'(s) \cdot \mathbf{n}(s)\,.$$

Essendo $\mathbf{b}(s) \cdot \mathbf{n}(s)$ identicamente zero, derivando si trova anche che

$$\tau(s) = \mathbf{b}(s) \cdot \mathbf{n}'(s)\,. \tag{6.19}$$

Il numero $\tau(s)$ può essere positivo negativo o nullo. Esso si chiama la *torsione* della curva.

Cerchiamo ora di esprimere $\mathbf{n}'(s)$ mediante $\mathbf{t}(s)$ e $\mathbf{b}(s)$. Ciò è possibile perchè $\mathbf{n}'(s)$, essendo ortogonale a $\mathbf{n}(s)$, è nel piano di $\mathbf{t}(s)$ e di $\mathbf{b}(s)$. Dunque

$$\mathbf{n}'(s) = \gamma(s)\mathbf{t}(s) + \delta(s)\mathbf{b}(s)\,. \tag{6.20}$$

Moltiplicando scalarmente i due membri di (6.20) per $\mathbf{t}(s)$ ed usando (6.17) si trova

$$\gamma(s) = -k(s)\,.$$

Analogamente, moltiplicando scalarmente (6.20) per $\mathbf{b}(s)$ ed usando (6.19) si trova

$$\delta(s) = \mathbf{n}'(s) \cdot \mathbf{b}(s) = \tau(s)\,.$$

Si trova quindi che i versori $\mathbf{t}(s)$, $\mathbf{n}(s)$, $\mathbf{b}(s)$ verificano

$$\begin{cases} \mathbf{t}'(s) &= k(s)\mathbf{n}(s) \\ \mathbf{n}'(s) &= -k(s)\mathbf{t}(s) + \tau(s)\mathbf{b}(s) \\ \mathbf{b}'(s) &= -\tau(s)\mathbf{n}(s)\,. \end{cases}$$

Si chiamano queste le *equazioni di Frenet* per curve di $\mathbf{R}^3$.

Il sistema di riferimento dato dai tre versori $\mathbf{t}(s)$, $\mathbf{n}(s)$, $\mathbf{b}(s)$ applicati in $\mathbf{r}(s)$ si chiama ancora il *riferimento* *mobile* sulla curva.

6.4.2 Appendice: Curve in $\mathbf{R}^n$

La maggior parte delle considerazioni che abbiamo svolto si estendono senza alcuna difficoltà a curve

$$\gamma : \qquad t \to \mathbf{r}(t)\,, \qquad \mathbf{r} \in \mathbf{R}^n\,.$$

Per esempio è ancora vero che una curva in $\mathbf{R}^n$ che è semplice è identificata dal suo sostegno a meno dell'orientazione; si definisce ancora la lunghezza dell'arco γ ponendo

$$L_\gamma = \int_a^b ||\dot{\mathbf{r}}(t)||\,\mathrm{d}t\,;$$

è quindi possibile definire il parametro d'arco.

La tangente e la normale si definiscono ancora nel medesimo modo come per le curve in $\mathbf{R}^3$. Non esiste invece una unica "binormale". Per completare il il riferimento mobile sulla curva si devono introdurre, oltre alla tangente ed alla normale, altri $n-2$ vettori e quindi le equazioni di Frenet diventano più complesse.

Capitolo 7

Integrazione delle funzioni di più variabili

In questo capitolo introdurremo gli integrali di funzioni di due o tre variabili. Le idee che si seguono per definire l'integrale sono simili a quelle che si usano per definire l'integrale delle funzioni di una sola variabile e quindi le illustreremo per sommi capi[1].

Una differenza importante tra le definizione di integrali per le funzioni di una o più variabili è nella scelta dei domini di integrazione. Per le funzioni di una variabile è naturale scegliere gli intervalli. Per le funzioni di più variabili c'è molta più libertà. Noi sceglieremo come domini di integrazione gli insiemi che sono delimitati da grafici di funzioni continue (si veda più avanti per una definizione più precisa). Tali insiemi si chiameranno domini di integrazione.

Studieremo la definizione dell'integrale di di una classe di funzioni che saranno costruite a partire da funzioni continue su insiemi chiusi e limitati.

[1] sul piano non si introduce una relazione d'ordine; e quindi non si introdurà per l'integrale multiplo un concetto analogo a quello di integrale orientato che si introduce nel caso delle funzioni di una variabile. Esiste però nel piano un "verso positivo di rotazione". Questo permetterà di introdurre "integrali orientati" sulle superfici in un capitolo successivo.

Dovremo usare la seguente proprietà delle funzioni che sono **continue su un qualsiasi insieme insieme chiuso e limitato** $\mathcal{D}$:

per ogni $\epsilon > 0$ esiste $\delta > 0$ con questa proprietà:
se R è un qualsiasi rettangolo contenuto in $\mathcal{D}$ le cui diagonali hanno lunghezza minore di δ allora

$$0 \leq \max_R f - \min_R f \leq \epsilon\,.$$

La proprietà importante è che **la posizione di R in $\mathcal{D}$ non ha importanza.** Inoltre, non abbiamo indicato esplicitamente le variabili della funzione perché questo risultato vale per funzioni di un qualsiasi numero di variabili (anche per funzioni di una sola variabile. In questo caso $R = [a, b]$, nonostante che questo risultato non sia stato provato nel corso di Analisi Matematica 1.)
Questa proprietà delle funzioni continue su insiemi chiusi e limitati si chiama *continuità uniforme.*

Introduciamo prima l'integrale di funzioni di due variabili, che presenteremo con maggiori dettagli, e poi l'integrale delle funzioni di tre variabili, che presenteremo in modo più conciso.

7.1 Integrazione delle funzioni di due variabili

Chiameremo *dominio di integrazione semplice* un insieme che ha una delle proprietà seguenti:

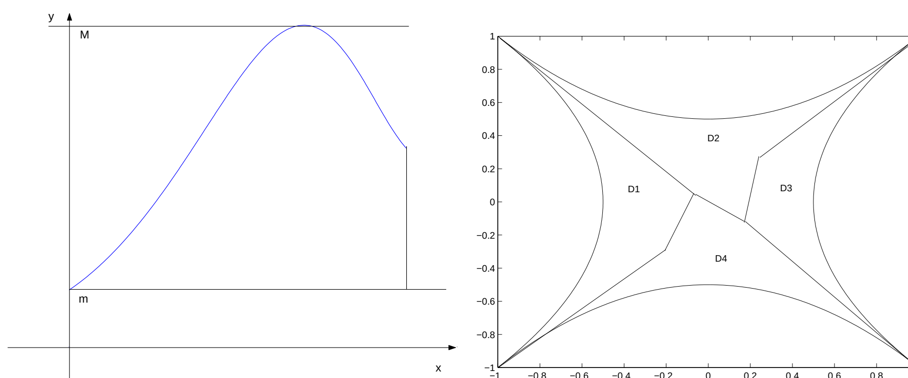
- è trapezoide di una funzione $y = g(x)$ continua definita su un intervallo limitato e chiuso (dell'asse delle ascisse) oppure $x = g(y)$ continua su un intervallo limitato e chiuso (dell'asse delle ordinate);
- oppure è differenza insiemistica di trapezoidi.

Va notato esplicitamente che ogni dominio di integrazione semplice è limitato, per il Teorema di Weiestrass. Inoltre, **assumiamo che il dominio di integrazione semplice sia chiuso, ossia contenga i punti della sua frontiera.** Nel caso specifico di un dominio di integrazione semplice compreso tra due grafici di funzioni della variabile $x \in [a, b]$, includeremo nel dominio sia i due grafici che i due segmenti che lo delimitano a destra (punti di ascissa a) ed a sinistra (punti di ascissa b).

Un insieme **chiuso e limitato** $\mathcal{D}$ si chiama *dominio di integrazione* quando si può rappresentare come unione di domini di integrazione semplici in modo tale che due qualsiasi di essi non abbiano punti interni comuni (ossia, gli eventuali punti comuni a due domini di integrazione semplici appartengono alle rette o ai grafici che li delimitano)[2].

[2]si faccia attenzione al fatto che "dominio" indica una insieme connesso ed aperto mentre i "domini di integrazione" sono insiemi chiusi.

Figura 7.1: Domini di integrazione



Per brevità un dominio di integrazione semplice si chiamerà "dominio semplice" ma è importante ricordare che è un caso speciale di dominio di integrazione.

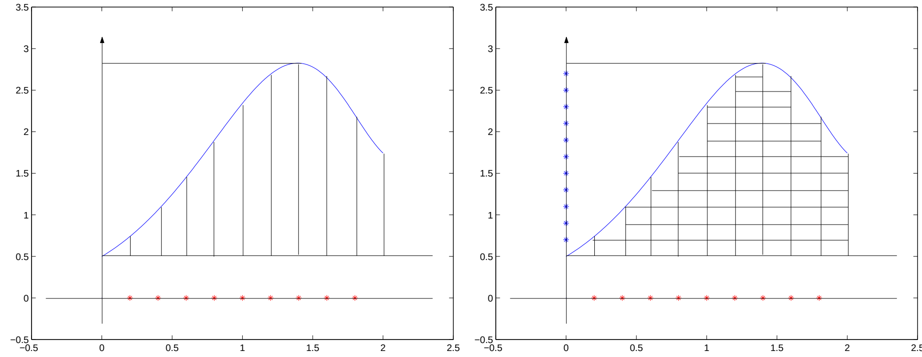
In conclusione, un dominio di integrazione si rappresenta (non in modo unico) come unione di domini semplici

$$\mathcal{D} = D_1 \cup D_2 \cup \cdots \cup D_n \quad D_i \cap D_j \text{ privo di punti interni.} \tag{7.1}$$

La figura 7.1 mostra esempi di domini di integrazione. Quello a sinistra è la parte del trapezoide di una funzione $y = g(x)$ $x \in [a, b]$ che è sopra alla retta orizzontale $y = m$ ed automaticamente sotto alla retta $y = M = \max g(x)$. E' quindi la differenza insiemistica tra il trapezoide di $g(x)$ e quello della funzione che vale costantemente m. Il dominio di integrazione a destra è più complesso, suddiviso mediante grafici di varie funzioni $y = y(x)$, oppure $x = x(y)$, ciascuna continua e definita su un opportuno intervallo limitato e chiuso.
Osservazione 204 Si noti che:

- un dominio di integrazione è un insieme chiuso e limitato. Questo implica che una funzione continua definita su un dominio di integrazione è limitata.
- Come abuso di linguaggio, se Ω è un dominio di integrazione, talvolta chiameremo dominio di integrazione l'insieme dei suoi punti interni; ossia l'insieme Ω privato delle curve che lo delimitano. Anche in questo caso, le funzioni continue che integreremo dovranno avere estenzione continua alla frontiera dell'insieme.
- uno stesso dominio di integrazione può rappresentarsi in più modi come unione di insiemi semplici: per esempio il trapezoide della funzione $\sin x$, $0 \le x \le \pi$ è un dominio di integrazione semplice che può rappresentarsi anche come $\mathcal{D}_1 \cup \mathcal{D}_2$ con $\mathcal{D}_1$ il trapezoide di $\sin x$, $0 \le x \le \pi/2$ e $\mathcal{D}_2$ quello di $\sin x$, $\pi/2 \le x \le \pi$.
- Si potrebbe provare che la chiusura di una regione di Jordan delimitata da una curva regolare è un dominio di integrazione; e quindi anche l'insieme ottenuto togliendogli la regione delimitata da una seconda curva di Jordan regolare lo è. ∎

Figura 7.2: Suddivisione in rettangoli di un dominio di integrazione



Introduciamo ora la classe di funzioni per cui definiremo l'integrale[3]. Chiamiamo *integrabile* una funzione $f(x,y)$ definita su un dominio di integrazione $\mathcal{D}$ (o almeno nei suoi punti interni) e con questa proprietà: E' possibile trovare una rappresentazione di $\mathcal{D}$ come in (7.1) in modo tale che per ogni indice i la funzione $f_{|D_i}(x,y)$ ammetta estensione continua al dominio semplice D_i. Ossia, **gli eventuali punti di discontinuità della funzione $f(x,y)$ devono essere sulle rette o grafici che delimitano i domini semplici che compongono $\mathcal{D}$.**

7.1.1 La definizione di integrale

Sia $f(x,y)$ una funzione integrabile. Per definirne l'integrale, procediamo in questo modo: definiamo l'integrale di $f(x,y)$ su ciascuna dei domini semplici D_i e quindi definiamo l'integrale su $\mathcal{D}$ come **somma** degli integrali sui domini D_i.

Si noti che il dominio $\mathcal{D}$ si potrà decomporre in più modi e quindi andrebbe provato che l'integrale di $f(x,y)$ non dipende dalla decomposizione scelta per il dominio. Questo è vero ma noi non lo proveremo.

Per semplicità limitiamoci a illustrare la definizione dell'integrale di $f(x,y)$ sul dominio semplice rappresentato nella figura 7.1 a sinistra. Indichiamo con T tale dominio semplice,

$$T = \{(x,y)\,, \quad 0 \le a \le b\,, \quad m \le g(x) \le M\}$$

Si segua il procedimento guardando la figura 7.2.

Dividiamo $[a,b]$ in N parti uguali con i punti $a_0 = a$, $a_k = k\frac{b-a}{N}$. Consideriamo quindi i segmenti verticali i cui punti hanno ascissa a_k e che sono contenuti in T, come in fig. 7.2, a sinistra.

Suddividiamo ora anche il segmento $[m,M]$ dell'asse delle ordinate in L tratti uguali. Le rette orizzontali i cui punti hanno ordinata uguale ai punti di suddivisione dell'intervallo

[3]si potrebbe definire l'integrale anche per funzioni con proprietà assai più generali, ma questa classe di funzioni è sufficiente praticamente per tutte le applicazioni.

$[m, M]$ disegnano una divisione dell'insieme T in rettangoli completamente contenuti in T, e inoltre **certi sottoinsiemi a lati non rettilinei, che escludiamo.** Si guardi la figura 7.2, a destra.

In questo modo il trapezoide della funzione viene ad essere approssimato da una rete di al più NL rettangolini, che indichiamo con $R_{i,j}$, $0 \le i \le N-1$, $0 \le j \le L-1$.

Ciascuno di questi rettangoli ha area $\frac{b-a}{N}\frac{M-m}{L}$. Indichiamo con U_- la loro unione.

Notiamo che nessuno dei rettangolini è a cavallo del grafico di $g(x)$ e che $U_- \subseteq T$.

Per ipotesi, la funzione $f(x, y)$ appartiene alla classe delle funzioni integrabili e quindi la sua restrizione a T ammette estensione continua a tutti i punti del dominio di integrazione T, che è chiuso e limitato. Indichiamo ancora con $f(x, y)$ tale estensione, che è uniformemente continua.

Costruiamo ora le somme

$$s_{N,L} = \sum_{i,j} \min_{(x,y)\in R_{i,j}} f(x,y) \cdot \frac{M-m}{L} \cdot \frac{b-a}{N}$$
$$S_{N,L} = \sum_{i,j} \max_{(x,y)\in R_{i,j}} f(x,y) \cdot \frac{M-m}{L} \cdot \frac{b-a}{N}$$

Queste somme sono estese a tutti i rettangoli che appartengono ad U_-.

Ora usiamo l'uniforme continuità di $f(x, y)$. Si fissi un qualsiasi $\epsilon > 0$ e il corrispondente $\delta > 0$. Se N ed L sono abbastanza grandi, diciamo $N \ge N_0$, $L \ge L_0$, la diagonale di ciascuno dei rettangoli costruiti misura meno di δ e quindi per $N \ge N_0$, $L \ge L_0$ si ha

$$0 \le [\max_{R_{i,j}} f(x,y) - \min_{R_{i,j}} f(x,y)] < \epsilon\,.$$

In particolare

$$0 \le S_{N_0,L_0} - s_{N_0,L_0} \le \{\text{somma delle aree dei rettangoli}\}\,\epsilon \le [(b-a)(M-m)]\,\epsilon\,.$$

Ma,

$$0 \le \inf\{S_{N,L}\} - \sup\{s_{N,L}\} \le S_{N_0,L_0} - s_{N_0,L_0} \le$$
$$\le \{\text{somma delle aree dei rettangoli}\}\,\epsilon \le [(b-a)(M-m)]\,\epsilon\,.$$

Dunque

$$\inf\{S_{N,L}\} = \sup\{s_{N,L}\}$$

e questo numero si chiama l' *integrale* di $f(x, y)$ sul trapezoide T. Esso si indica col simbolo

$$\int_T f(x,y)\,\mathrm{d}x\ \mathrm{d}y\,.$$

In modo analogo si definisce l'integrale su ogno altro dominio D_i. Si definisce quindi

$$\int_{\mathcal{D}} f(x,y)\,\mathrm{d}x\ \mathrm{d}y = \sum_i \int_{D_i} f(x,y)\,\mathrm{d}x\ \mathrm{d}y\,. \qquad (7.2)$$

Osservazione 205 E' ovvio dalla costruzione che abbiamo fatto che se $f(x, y) \ge 0$ allora il suo integrale si interpreta come il *volume* del solido compreso tra l'insieme $\mathcal{D}$ del piano

$z = 0$ ed il grafico della funzione. Se in particolare si sceglie $f(x, y)$ identicamente uguale ad 1, si trova una numero che ha senso interpretare come *area* del dominio di integrazione. Ossia **per definizione:**

$$\text{area di } \mathcal{D} = \int_{\mathcal{D}} 1 \, \mathrm{d}x \, \mathrm{d}y \,. \tag{7.3}$$

Se accade che $\mathcal{D}$ è grafico di una funzione $y = g(x) \geq 0$ continua su $[a, b]$ abbiamo due **diverse** definizioni di area: quella appena scritta e quella data durante il corso di Analisi Matematica 1: area del trapezoide di una funzione positiva uguale (per definizione!) al suo integrale. Si tratta di due definizioni diverse ma vedremo che conducono al medesimo numero. **Un fatto da sottolineare è questo: $\int_a^b g(x) \, \mathrm{d}x$ è l'area del trapezoide solo se $g(x) \geq 0$. Invece, la (7.3) dà il valore dell'area anche quando $\mathcal{D}$ è trapezoide di una funzione che cambia segno.** ∎

Talvolta, un integrale di una funzione di due variabili si chiama anche *integrale doppio*. Per contrasto, l'integrale di una funzione di una sola variabile,

$$\int_a^b f(x) \, \mathrm{d}x$$

si chiama anche *integrale semplice*.

7.1.2 Le proprietà dell'integrale

Le proprietà dell'integrale sono le stesse come nel caso degli integrali semplici:

- la linearità: se α e β sono numeri e $f(x, y)$, $g(x, y)$ sono funzioni continue sullo stesso dominio di integrazione $\mathcal{D}$, vale

$$\int_{\mathcal{D}} [\alpha f(x, y) + \beta g(x, y)] \, \mathrm{d}x \, \mathrm{d}y = \alpha \int_{\mathcal{D}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y + \beta \int_{\mathcal{D}} g(x, y) \, \mathrm{d}x \, \mathrm{d}y \,;$$

- additività: se il dominio di integrazione $\mathcal{D}$ è unione di due,

$$\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$$

  e se la funzione $f(x, y)$ è integrabile sia su $\mathcal{D}_1$ che su $\mathcal{D}_2$, allora è anche integrabile su $\mathcal{D}$, e viceversa; e inoltre

$$\int_{\mathcal{D}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{\mathcal{D}_1} f(x, y) \, \mathrm{d}x \, \mathrm{d}y + \int_{\mathcal{D}_2} f(x, y) \, \mathrm{d}x \, \mathrm{d}y \,.$$

- monotonia: $f(x, y) \leq g(x, y)$ per ogni $(x, y) \in \mathcal{D}$ implica

$$\int_{\mathcal{D}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y \leq \int_{\mathcal{D}} g(x, y) \, \mathrm{d}x \, \mathrm{d}y \,.$$

- Dalla monotonia si deduce

$$\left| \int_{\mathcal{D}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y \right| \leq \int_{\mathcal{D}} |f(x, y)| \, \mathrm{d}x \, \mathrm{d}y \,.$$

Anche il *teorema della* *media* si può riformulare. Indichiamo con $A(\mathcal{D})$ l'area di $\mathcal{D}$, calcolata sommando le aree delle singole regioni che la compongono.

Vale:

$$A(\mathcal{D}) \cdot \left(\min_{(x,y)\in\mathcal{D}} f(x,y) \right) \leq \int_{\mathcal{D}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \leq A(\mathcal{D}) \cdot \left(\max_{(x,y)\in\mathcal{D}} f(x,y) \right) .$$

Ne segue:

Teorema 206 *Se $\mathcal{D}_\epsilon$ è una circonferenza di raggio ϵ e centro fissato vale*

$$\lim_{\epsilon\to 0} \int_{\mathcal{D}_\epsilon} f(x,y)\,dx\;dy = 0\,.$$

7.1.3 Domini di integrazione definiti mediante curve di Jordan

Sia

$$\gamma\,:\,t \to x(t)\mathbf{i} + y(t)\mathbf{j}$$

una curva di Jordan regolare. Applicando il *Teorema della funzione implicita* si vede che il sostegno è localmente un grafico e si potrebbe provare che il sostegno è grafico di un numero *finito* di funzioni. Dunque, l'unione della regione Ω_γ, interna alla curva γ e del suo sostegno è un dominio di integrazione. Come si è notato all'osservazione 204, diremo che la regione Ω_γ stessa è un dominio di integrazione e quindi è possibile:

- integrare una funzione su una regione di Jordan, se la funzione è continua sulla chiusura della regione;
- assegnare un'area ad una regione di Jordan (ciò che si fa integrando la funzione identicamente uguale ad 1);
- approssimare l'area di una regione di Jordan mediante la somma delle aree di un numero finito di rettangoli con lati paralleli agli assi coordinati.

Con lo stesso abuso di linguaggio illustrato all'osservazione 204, diremo ancora che una regione di Jordan si può approssimare mediante rettangoli.

7.1.4 Riduzione di integrali doppi ad integrali iterati

Torniamo a considerare le somme $s_{N,L}$ ed $S_{N,L}$ che servono per definire l'integrale doppio. Consideriamo per esempio le $s_{N,L}$:

$$s_{N,L} = \sum_{i,j} \left[\min_{(x,y)\in R_{i,j}} f(x,y) \right] \cdot \frac{M-m}{L} \cdot \frac{b-a}{N}\,.$$

Calcoliamo le somme prima di tutto sommando i termini che corrispondono a rettangolini che appartengono alla stessa striscia verticale, ossia scrivendo

$$\begin{aligned} s_{N,L} &= \sum_{i,j} \left[\min_{(x,y)\in R_{i,j}} f(x,y) \right] \cdot \frac{M-m}{L} \cdot \frac{b-a}{N} \\ &= \sum_{i} \left\{ \sum_{j} \left[\min_{(x,y)\in R_{i,j}} f(x,y) \right] \cdot \frac{M-m}{L} \right\} \cdot \frac{b-a}{N}\,. \end{aligned}$$

Fissiamo un punto x_i in ciascuno degli intervalli $[a_i, a_{i+1})$. Si mostra che per $L \to +\infty$ tende a zero la differenza tra la parentesi graffa e l'integrale

$$\int_0^{g(x_i)} f(x_i, y)\, \mathrm{d}y$$

ove x_i è un qualsiasi punto dell'intervallo $[a_i, a_{i+1})$; e quindi che

$$\begin{aligned} s_{N,L} &= \sum_{i,j} \left[\min_{(x,y)\in R_{i,j}} f(x,y) \right] \cdot \frac{M-m}{L} \cdot \frac{b-a}{N} \\ &= \sum_i \frac{b-a}{N} \int_0^{g(x_i)} f(x_i, y)\, \mathrm{d}y + \epsilon(L, N) \end{aligned}$$

con

$$\lim \epsilon(L, N) = 0\,;$$

Ma, per $N \to +\infty$, le somme

$$\sum_{i=0}^{N-1} \frac{b-a}{N} \int_0^{g(x_i)} f(x_i, y)\, \mathrm{d}y$$

convergono all'integrale della funzione di x

$$\int_0^{g(x)} f(x, y)\, \mathrm{d}y\,,$$

ossia all'integrale iterato di $f(x, y)$. Dunque, per calcolare $\int_{\mathcal{D}} f(x, y)\, \mathrm{d}x\ \mathrm{d}y$ si può procedere come segue:

1. Si proietta ortogonalmente $\mathcal{D}$ sull'asse delle ascisse, ottenendo un intervallo $[a, b]$;
2. Si traccia la retta parallela all'asse delle ordinate e che passa da $x \in [a, b]$. Si indica con $\mathcal{S}_x$ l'intersezione di tale retta con $\mathcal{D}$. L'insieme $\mathcal{S}_x$ è unione di un numero finito di intervalli.
3. Si ha: $\int_{\mathcal{D}} f(x, y)\, \mathrm{d}x\ \mathrm{d}y = \int_a^b \left[\int_{\mathcal{S}_x} f(x, y)\, \mathrm{d}y \right] \mathrm{d}x\,.$

Si veda la figura 7.3

Naturalmente la stessa procedura vale anche scambiando il ruolo dell'asse delle ascisse con quello dell'asse delle ordinate.

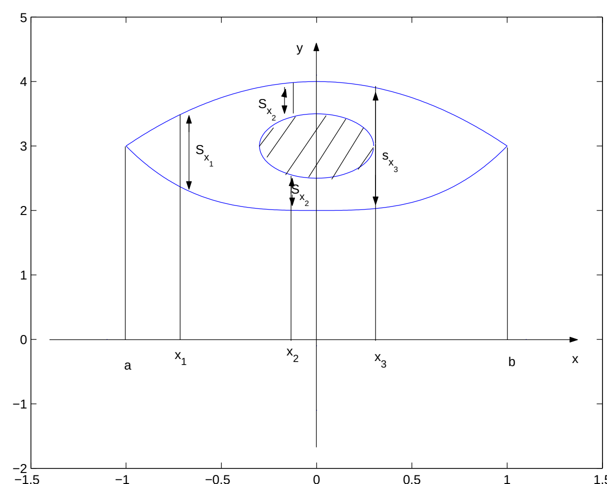
Consideriamo un caso particolare: supponiamo che $\mathcal{D}$ sia il trapezoide della funzione $k(x)$, $x \in [a, b]$, e che la funzione integranda sia identicamente uguale ad 1. Sia inoltre $k(x)$ non negativa. In tal caso,

$$\int_{\mathcal{D}} 1\ \mathrm{d}x\ \mathrm{d}y = \int_a^b \left[\int_0^{k(x)} 1\ \mathrm{d}y \right] \mathrm{d}x = \int_a^b k(x)\, \mathrm{d}x$$

in accordo con quanto detto nell'osservazione 205.

Il metodo visto "riduce" il calcolo di un integrale doppio a quello di un integrale iterato, e quindi a quello di due integrali semplici. Però esso può anche usarsi al contrario, per ricondurre il calcolo di un integrale iterato calcolato prima rispetto ad x e poi rispetto ad y al calcolo di un integrale doppio; e quindi al calcolo di un integrale iterato calcolato prima rispetto ad y e poi rispetto ad x. Quando si opera in questo modo su un integrale iterato si dice che si *scambia l' ordine d'integrazione*.

Figura 7.3: Riduzione di un integrale doppio



7.2 Integrazione delle funzioni di tre variabili

L'integrazione delle funzioni di tre variabili si introduce in modo del tutto analogo a quella relativa a funzioni di due variabili. Prima di tutto si scelgono i *domini di integrazione semplici:* questi sono i solidi delimitati dai grafici di due funzioni, per esempio

$$\phi(x, y) \leq z \leq \psi(x, y)$$

con $(x, y) \in \mathcal{D}$, dove $\mathcal{D}$ è un dominio di integrazione semplice per funzioni di due variabili.

Anche i *domini di integrazioni semplici* per funzioni di tre variabili sono insiemi chiusi.

Come *dominio di integrazione* per funzioni di tre variabili intendiamo l'unione di un numero finito di tali domini semplici, purché l'intersezione tra due qualsiasi di essi non contenga punti interni.

Definiamo ora l'insieme delle *funzioni integrabili.* Una funzione definita su un dominio di integrazione si dice integrabile quando la sua restizione all'interno di ciascuno dei domini semplici ammette estensione continua al dominio semplice stesso.

Ciò detto è facile dividere un dominio di integrazione in "piccoli" parallelepipedi e costruire le analoghe delle somme s_N ed S_N e quindi definire

$$\int_{\mathcal{D}} f(x, y, z)\, dx\ dy\ dz$$

come estremo comune ai due insiemi $\{s_{N,L,K}\}$ ed $\{S_{N,L,K}\}$ che ora verranno a dipendere da tre indici N, L, K.

Si ottiene così un integrale che si chiama anche *integrale triplo*.

Osservazione 207 Per definizione, il *volume* di un dominio di integrazione $\mathcal{D} \subseteq \mathbf{R}^3$ è

$$\int_{\mathcal{D}} 1 \, \mathrm{d}x \; \mathrm{d}y \; \mathrm{d}z \,. \quad \blacksquare$$

Le proprietà elencate al paragrafo 7.1.2 per gli integrali doppi valgono anche per gli integrali tripli. Per gli integrali tripli, il Teorema 210 si riformula come segue:

Teorema 208 *Sia $(u,v,w) \mapsto (x(u,v,w), y(u,v,w), z(u,v,w))$ una trasformazione continua da un dominio di integrazione $\mathcal{D}_1$ ad un dominio di integrazione $\mathcal{D}$. Supponiamo che la trasformazione ammetta derivate parziali continue nei punti interni di $\mathcal{D}_1$ e che le derivate abbiano estensione continua alla frontiera, e che il suo determinante jacobiano $J(u,v,w)$ non si annulli.*

Per ogni funzione $f(x,y,z)$ continua su $\mathcal{D}$ si ha:

$$\int_{\mathcal{D}} f(x,y,z) \, dx \; dy \; dz = \int_{\mathcal{D}_1} f\left(x(u,v,w), y(u,v,w), z(u,v,w)\right) J(u,v,w) \, du \; dv \; dw \,.$$

Invece, il metodo di riduzione va esaminato esplicitamente.

7.3 Formula di riduzione per gli integrali tripli

Il calcolo degli integrali tripli si può ricondurre al calcolo di integrali iterati. Illustriamo il metodo nel caso particolare in cui il dominio di integrazione $\mathcal{D}$ è compreso tra due grafici

$$\mathcal{D} = \{(x,y,z) \mid \phi(x,y) \leq z \leq \psi(x,y)\} \,.$$

Indichiamo con

$$\mathcal{D}_z = \{(x,y) \mid \exists w \text{ per cui } (x,y,w) \in \mathcal{D}\}$$

ossia la proiezione ortogonale di $\mathcal{D}$ sul piano $z = 0$. Dunque:

- se $(x,y) \notin \mathcal{D}_z$ allora la retta verticale per (x,y) non interseca $\mathcal{D}$;
- se $(x,y) \in \mathcal{D}_z$ allora la retta verticale per (x,y) interseca $\mathcal{D}$ nel segmento verticale di estremi $(x,y,\phi(x,y))$ e $(x,y,\psi(x,y))$. Si noti che questo segmento potrebbe essere ridotto ad un punto.

Vale:

$$\int_{\mathcal{D}} f(x,y,z) \, \mathrm{d}x \; \mathrm{d}y \; \mathrm{d}z = \int_{\mathcal{D}_z} \left[\int_{\phi(x,y)}^{\psi(x,y)} f(x,y,z) \, \mathrm{d}z \right] \mathrm{d}x \; \mathrm{d}y \,.$$

In questo modo il calcolo dell'integrale triplo si è ricondotto al calcolo di un integrale semplice, seguito da quello di un integrale doppio[4].

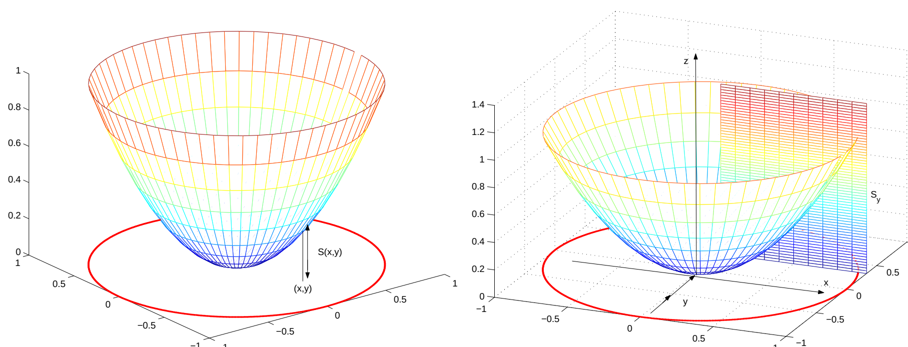
Si veda la figura 7.4, a sinistra.

Si può anche procedere in modo diverso: supponiamo di sapere che $\mathcal{D}_z$ sia delimitato da due grafici, per esempio

$$\mathcal{D}_z = \{(x,y) \mid h(y) \leq x \leq k(y)\}$$

[4]questo metodo di riduzione si chiama anche "metodo di riduzione per fili".

Figura 7.4: Riduzione per fili e per strati di un integrale triplo



e sia $[a, b]$ il dominio comune ad $h(y)$ e $k(y)$. Allora vale:

$$\int_{\mathcal{D}} f(x, y, z)\, \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = \int_a^b \left[\int_{\tilde{\mathcal{D}}_y} f(x, y, z)\, \mathrm{d}x\ \mathrm{d}z \right] \mathrm{d}y$$

dove $\tilde{\mathcal{D}}_y$ è l'intersezione di $\mathcal{D}$ col piano parallelo agli assi x e z, passante per il punto $(0, y, 0)$[5]. Si veda la figura 7.4, a destra.

7.3.1 Integrazione e Cambiamento di variabili

Nel caso degli integrali semplici, sotto opportune ipotesi si prova la formula

$$\int_a^b f(x)\, \mathrm{d}x = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(t))\phi'(t)\, \mathrm{d}t\,.$$

Si noti però che $\phi^{-1}(a)$ può anche essere maggiore di $\phi^{-1}(b)$, ciò che è lecito perché nel caso degli integrali semplici abbiamo definito l'integrale orientato. D'altra parte, questa formula non si prova usando direttamente le proprietà dell'integrale. Piuttosto si prova che questa formula vale per il calcolo delle primitive, e quindi anche per il calcolo dell'integrale grazie al teorema fondamentale del calcolo integrale. Mostriamo come a questa formula, sotto opportune ipotesi, si possa anche giungere direttamente dalla definizione di integrale semplice.

Sia $f(x)$ una funzione continua definita su in intervallo $[a, b]$ e sia $x = \phi(t)$ una funzione monotona strettamente crescente da un intervallo $[\alpha, \beta]$ su $[a, b]$, che è anche derivabile.

Bisogna ricordare questi fatti:

- nella definizione di integrale la suddivisione dell'intervallo $[a, b]$ non è necessariamente fatta mediante punti equidistanti;

[5]questo metodo di riduzione si chiama anche "metodo di riduzione per strati".

- nella definizione di integrale, si possono considerare le somme

$$\sum_{i=1}^{n} f(\xi_i)[a_{i+1} - a_i] \tag{7.4}$$

con punti $\xi_i \in (a_i, a_{i+i})$ qualsiasi.

Quando la finezza della partizione tende a zero, le somme (7.4) tendono all'integrale

$$\int_a^b f(x)\,\mathrm{d}x\,.$$

Essendo la funzione ϕ monotona e suriettiva, ogni a_i proviene da un unico α_i,

$$a_i = \phi(\alpha_i).$$

Dunque la somma (7.4) è anche uguale a

$$\sum_{i=1}^{n} f(\xi_i)[\phi(\alpha_{i+1}) - \phi(\alpha_i)]\,.$$

Questa non è una somma di quelle che conducono alla definizione di un integrale, ma ad essa facilmente si riconduce. Infatti, dal Teorema di Lagrange, esiste un punto c_i tale che

$$[\phi(\alpha_{i+1}) - \phi(\alpha_i)] = \phi'(c_i)(\alpha_{i+1} - \alpha_i).$$

Ricordando che i numeri ξ_i si possono scegliere in modo arbitrario, scegliamo $\xi_i = \phi(c_i)$. La monotonia di ϕ mostra che $c_i \in (\alpha_i, \alpha_{i+1})$. In questo modo la (7.4) diviene

$$\sum_{i=1}^{n} f(\phi(c_i))\phi'(c_i)[\alpha_{i+1} - \alpha_i]\,. \tag{7.5}$$

La finezza della partizione di $[a, b]$ tende a zero se e solo se tende a zero la finezza della partizione di $[\alpha, \beta]$. Quando la finezza della partizione tende a zero, le somme (7.4) tendono all'integrale di $f(x)$, quelle di (7.5) tendono all'integrale di $f(\phi(t))\phi'(t)$. D'altra parte le due somme hanno lo stesso valore e quindi si trova

$$\int_a^b f(x)\,\mathrm{d}x = \int_\alpha^\beta f(\phi(t))\phi'(t)\,\mathrm{d}t\,.$$

Osservazione 209 Notiamo esplicitamente:

- in questo calcolo la crescenza di ϕ si è usata: è $\alpha_i < \alpha_{i+1}$ proprio perchè ϕ è crescente. Se invece ϕ decresce, sarà $\beta < \alpha$ e in (7.5) si ha $\alpha_{i+1} < \alpha_i$ e quindi al limite si trova

$$\int_a^b f(x)\,\mathrm{d}x = \int_\alpha^\beta f(\phi(t))\phi'(t)\,\mathrm{d}t\,.$$

Ma ora $\beta < \alpha$ e quindi riordinando l'ordine degli estremi di integrazione si trova

$$\int_a^b f(x)\,\mathrm{d}x = -\int_\beta^\alpha f(\phi(t))\phi'(t)\,\mathrm{d}t$$
$$= \int_\beta^\alpha f(\phi(t))[-\phi'(t)]\,\mathrm{d}t = \int_\beta^\alpha f(\phi(t))|\phi'(t)|\,\mathrm{d}t\,.$$

- il ruolo del numero $\phi'(c)$: è il coefficiente che trasforma la lunghezza di $[\alpha_i, \alpha_{i+1}]$ nella lunghezza di $[a_i, a_{i+1}]$. Si noti che se ϕ' non si annulla su $[a, b]$ (estremi inclusi) esistono numeri m, M tali che

$$m|\alpha_i - \alpha_{i+1}| \leq |a_i - a_{i+1}| \leq M|\alpha_i - \alpha_{i+1}|\,. \quad \blacksquare$$

- Nel caso degli integrali semplici, la formula vale anche senza richiedere la monotonia di ϕ perchè, intuitivamente, se per t che percorre $[\alpha, \beta]$ il punto $\phi(t)$ percorre più volte un intervallo $[x', x''] \subseteq [a, b]$, deve percorrerlo un numero dispari di volte, in versi opposti; e grazie all'esistenza dell'integrale orientato, i contributi dei passaggi 2 e 3 si elidono, lo stesso per i passaggi 4 e 5 ecc. Un fenomeno analogo non potrà aversi per funzioni di più variabili e quindi in tal caso dovremo imporre a ϕ di essere biunivoca. ∎

Vediamo ora quali problemi si incontrano nel cercare di estendere il ragionamento appena fatto a funzioni di più variabili. In questo caso dovremo avere $f(x, y)$ definita su un dominio di integrazione $\mathcal{D}$ e dovremo avere una trasformazione $(x, y) = \Phi(u, v) = (\phi_1(u, v), \phi_2(u, v))$ da un dominio di integrazione $\mathcal{D}'$ nel dominio di integrazione $\mathcal{T}$.

La Φ dovrà essere biunivoca da $\mathcal{D}'$ su $\mathcal{D}$ e differenziabile (un'ulteriore condizione si dirà in seguito).

L'integrale di $f(x, y)$ si definisce suddividendo $\mathcal{D}$ in tanti piccoli rettangoli, diciamo R_i. Un rettangolo R_i è immagine mediante Φ di un sottoinsieme R'_i di $\mathcal{D}'$ che però non è un rettangolo. Se la Φ è lineare R' è un parallelogramma, altrimenti è una figura più complessa. Ciò nonostante, si potrà tentare di ripetere gli argomenti visti sopra se:

- si potrà trovare una relazione tra l'area di R_i e quella di R'_i;
- l'area dei rettangoli R_i tende a zero se e solo se l'area degli insiemi R'_i tende a zero.

La relazione tra l'area di R_i e quella di R'_i è nota nel caso in cui la trasformazione Φ sia lineare: è

$$(\text{area di } R_i) = |\det \Phi|(\text{area di } R'_i)\,.$$

Qui Φ indica la matrice della trasformazione Φ, calcolata rispetto a coordinate ortogonali.

Nel caso non lineare una formula analoga alla precedente ancora vale, con un errore che è di ordine superiore rispetto all'area di R', quando questa tende a zero.

Questo risponde alla prima questione. La seconda richiesta, l'area degli R_i tende a zero se e solo se l'area degli R'_i tende a zero, è soddisfatta quando $\det \Phi \neq 0$ su $\mathcal{D}'$.

Con queste informazioni, nel caso in cui Φ sia una trasformazione lineare è relativamente facile provare il risultato seguente, mimando la dimostrazione vista sopra per il caso di funzioni di una variabile. Nel caso in cui Φ sia non lineare, il risultato seguente vale ancora ma la dimostrazione è piuttosto complessa:

Teorema 210 *Siano $\mathcal{D}$ e $\mathcal{D}'$ due domini di integrazione. Sia*

$$(x, y) = \Phi(u, v) = (\phi_1(u, v), \phi_2(u, v))$$

una trasformazione invertibile *da $\mathcal{D}'$ su $\mathcal{D}$. Supponiamo che questa trasformazione sia di classe C^1 su una regione Ω che contiene $\mathcal{D}'$.*

Sia $\det J(u, v)$ lo jacobiano della trasformazione. Supponiamo che $\det J(u, v)$ non si annulli su $\mathcal{D}'$. *Sotto queste condizioni vale:*

$$\int_{\mathcal{D}} f(x, y)\, dx\ dy = \int_{\mathcal{D}'} f(\phi_1(u, v), \phi_2(u, v)) \cdot |\det J(u, v)|\ du\ dv\,.$$

Come abbiamo detto, questo teorema estende il teorema di cambiamento di variabili negli integrali semplici. Nell'uso però esso ha un ruolo diverso. Nel caso degli integrali semplici il metodo di cambiamento di variabili si usa per trasformare la funzione in una di cui sia più facile trovare la primitiva. Nel caso degli integrali doppi si usa la trasformazione di variabili per passare da un dominio più complicato ad uno più semplice. Per esempio, si voglia calcolare

$$\int_{\mathcal{D}} \sqrt{x^2+y^2}\,\mathrm{d}x\ \mathrm{d}y$$

con $\mathcal{D}$ la circonferanza $x^2+y^2 \leq 1$. Esprimendo x ed y mediante le coordinate polari ρ e θ, si trova

$$x=\rho\cos\theta\,, \qquad y=\rho\sin\theta\,, \qquad 0\leq\rho\leq 1\,, \qquad 0\leq\theta\leq 2\pi\,.$$

Notando che lo jacobiano della trasformazione è semplicemente ρ, il calcolo richiesto si riduce a quello dell'integrale iterato

$$\int_0^{2\pi}\left[\int_0^1 \rho^2\,\mathrm{d}\rho\right]\mathrm{d}\theta=\frac{2}{3}\pi\,.$$

Osservazione 211 Si noti che lo jacobiano della trasformazione a coordinate polari si annulla nell'origine e quindi il Teorema 210 a rigore non può applicarsi. Si applichi però il teorema ad una corona circolare $\epsilon \leq \rho \leq 1$ e poi si mandi ϵ a zero. Il Teorema 206 mostra che il contributo della circonferenza di raggio ϵ tende a zero e ciò giustifica l'uso delle coordinate polari per il calcolo precedente. ∎

7.4 Alcuni jacobiani che è importante ricordare

Le trasformazioni di coordinate che si usano più comunemente sono le trasformazioni a coordinate polari o ellittiche nel piano, a coordinate cilindriche o sferiche nello spazio.

Si ha:

coordinate	trasformazione	jacobiano
	Nel piano	
polari (ρ, θ)	$\begin{cases} x = \rho\cos\theta \\ y = \rho\sin\theta \end{cases}$	ρ
ellittiche (ρ, θ)	$\begin{cases} x = a\rho\cos\theta \\ y = b\rho\sin\theta \end{cases}$	$ab\rho$
	Nello spazio	
cilindriche (ρ, θ, z)	$\begin{cases} x = \rho\cos\theta \\ y = \rho\sin\theta \\ z = z \end{cases}$	ρ
sferiche (ρ, θ, ϕ)	$\begin{cases} x = \rho\cos\theta\sin\phi \\ y = \rho\sin\theta\sin\phi \\ z = \rho\cos\phi \end{cases}$	$\rho^2\sin\phi$

Si noti che nella formula di cambiamento di coordinate per gli integrali multipli compare il valore assoluto dello jacobiano, mentre la tavola precedente riporta lo jacobiano per sottolineare che i sistemi di coordinate che abbiamo introdotto, con le coordinate che si susseguono nell'ordine indicato nella definizione delle coordinate, hanno jacobiano positivo. Dunque la matrice jacobiana di tali trasformazioni non altera l'orientazione di $\mathbf{R}^3$.

7.4.1 Volumi delimitati da superfici di rotazione

Consideriamo il grafico di una funzione sul piano (y, z), descritto dall'equazione $z = f(y)$.

Per fissare le idee, supponiamo che la funzione sia definita (e continua) su $[0, Y]$ e che prenda valori positivi.

Facendo ruotare i punti del grafico intorno all'asse z, si trova la superficie descritta dall'equazione

$$z = f\left(\sqrt{x^2 + y^2}\right) .$$

Si vuol calcolare il volume dell'insieme

$$V = \left\{(x, y, z) \mid 0 \le z \le f\left(\sqrt{x^2 + y^2}\right)\right\} ;$$

ossia il volume dell'insieme compreso tra il piano $z = 0$ e la superficie.

Vogliamo quindi calcolare

$$\int_V \mathrm{d}x \; \mathrm{d}y \; \mathrm{d}z \, .$$

Passiamo a coordinate cilindriche

$$\begin{cases} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z \, . \end{cases}$$

Si calcola facilmente che lo jacobiano della trasformazione è r. Dunque va calcolato

$$\int_{\mathcal{R}} r \, \mathrm{d}r \; \mathrm{d}z \; \mathrm{d}\theta .$$

L'insieme $\mathcal{R}$ è ora

$$\mathcal{R} = \{(r, \theta, z) \mid 0 \le \theta \le 2\pi \, , \; 0 \le r \le Y \, , \; 0 \le z \le f(r)\} \; .$$

Dunque, riducendo l'integrale triplo ad integrali iterati si trova

$$V = \int_0^{2\pi} \left[\int_0^Y \left[\int_0^{f(r)} 1 \cdot r \, \mathrm{d}z \right] \mathrm{d}r \right] \mathrm{d}\theta = 2\pi \int_0^Y r f(r) \, \mathrm{d}r \, . \tag{7.6}$$

Supponiamo invece che il grafico che si fa ruotare sia dato mediante una funzione $y = f(z)$, $0 \le z \le Z$.

Passando a coordinate cilindriche

$$\begin{cases} x &= r \cos\theta \\ y &= r \sin\theta \\ z &= z \, , \end{cases}$$

si deve integrare ora sull'insieme

$$\tilde{\mathcal{R}} = \{(r, \theta, z) \mid 0 \le \theta \le 2\pi \, , \quad 0 \le z \le Z \, , \quad 0 \le r \le f(z)\} \; .$$

Dunque,

$$V = \int_V 1 \, \mathrm{d}x \; \mathrm{d}y \; \mathrm{d}z = \int_0^{2\pi} \left[\int_0^z \left(\int_0^{f(z)} 1 \cdot r \, \mathrm{d}r \right) \mathrm{d}z \right] \mathrm{d}\theta = \pi \int_0^Z f^2(z) \, \mathrm{d}z \, . \tag{7.7}$$

Questa formula si chiama *Formula di Guldino*.

L'interpretazione geometrica di queste formule si capisce facilmente approssimando gli integrali semplici mediante le somme di Riemann, che conducono alla definizione stessa degli integrali. Consideriamo prima di tutto l'integrale (7.6). Consideriamo una partizione dell'intervallo $[0, Y]$ mediante i punti y_i. Le somme di Riemann inferiori hanno forma

$$2\pi \left[\sum r_i f(r_i)(y_{i+1} - y_i) \right]$$

ove r_i è un punto di minimo della funzione $r f(r)$ nell'intervallo $[y_{i+1}, y_i]$.

Il punto $(0, r_i, 0)$ dell'assse delle ordinate ruota, per descrivere la superficie, sulla circonferenza di raggio r_i. Dunque, $2\pi r_i(y_{i+1} - y_i)$ è circa l'area della corona circolare descritta, durante la rotazione, dal segmento $[y_{i+1}, y_i]$. Dunque, $2\pi r_i(y_{i+1} - y_i) f(r_i)$ è (circa) il volume del più alto "guscio cilindrico" che insiste su tale corona circolare, e che è sotto al grafico della funzione.

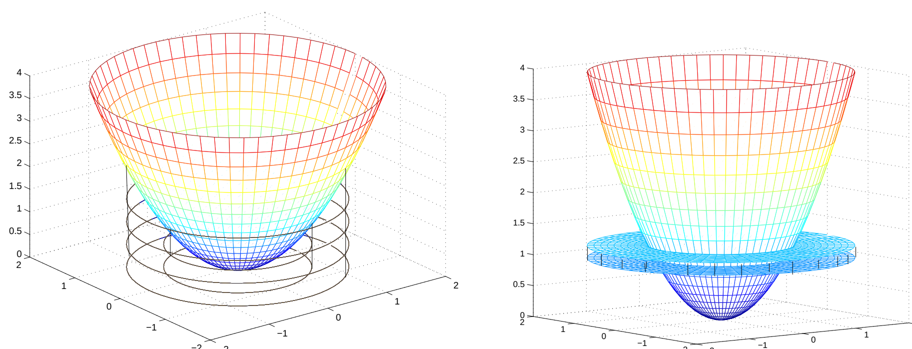
Invece, le somme di Riemann dell'integrale (7.7) si ottengono dividendo il segmento $[0, Z]$ con i punti z_i. Le somme di Riemann sono

$$\pi \left[\sum f^2(\tilde{z}_i)(z_{i+1} - z_i) \right] \, , \qquad \tilde{z}_i \in [z_i, z_{i+1}] \, .$$

Ciascun addendo $\pi f^2(\tilde{z}_i)(z_{i+1} - z_i)$ rappresenta circa il volume di un cilindretto di raggio $f(\tilde{z}_i)$ e base sul segmento $[z_{i+1}, z_i]$: in questo caso il volume si approssima come somma dei volumi di tali cilindretti.

La figura 7.5 illustra a sinistra il primo ed a destra il secondo caso (il volume da calcolare è quello del solido compreso tra la parte di paraboloide disegnata ed il piano $z = 0$).

Figura 7.5: Volume di una superficie di rotazione



7.5 Appendici

7.5.1 Appendice: Integrali impropri

Nel caso degli integrali di una variabile, è stato naturale definire

$$\int_0^{+\infty} f(x)\,\mathrm{d}x = \lim_{R\to+\infty} \int_0^R f(x)\,\mathrm{d}x\,.$$

Una definizione analoga nel caso di integrali doppi è è molto più delicata. Per vedere la ragione, consideriamo una funzione $f(x,y)$ definita su un insieme illimitato I, che interseca ogni disco

$$D_R = \{(x,y) \mid x^2+y^2 \le R^2\}$$

in un dominio di integrazione. Allora, si può definire

$$\int_I f(x,y)\,\mathrm{d}x\ \mathrm{d}y = \lim_{R\to+\infty} \int_{I\cap D_R} f(x,y)\,\mathrm{d}x\ \mathrm{d}y\,.$$

Però questa definizione privilegia i dischi. Niente garantisce che si giunga al medesimo limite se, invece di dischi, si considerano i quadrati

$$Q_R = \{(x,y) \mid |x| \le R\,,\ |y| \le R\}\,.$$

L'esempio seguente mostra che in generale usando dischi od usando quadrati si trovano comportamenti diversi.

Esempio 212 L'insieme I è il primo quadrante e la funzione che si considera è

$$f(x,y) = \sin\left(x^2+y^2\right)\,.$$

L'integrale si D_R si calcola immediatamente passando a coordinate polari:

$$\int_{D_R} f(x,y)\,\mathrm{d}x\ \mathrm{d}y = \frac{\pi}{4}\left(1-\cos R^2\right)\,,$$

privo di limite per $R \to +\infty$.

L'integrale su Q_R si calcola in coordinate cartesiane:

$$\int_{Q_R} f(x,y)\,\mathrm{d}x\ \mathrm{d}y = \int_{Q_R} (\sin x^2 \cos y^2 + \cos x^2 \sin y^2)\,\mathrm{d}x\ \mathrm{d}y$$
$$= \left[\int_0^R \sin x^2\,\mathrm{d}x\right]\left[\int_0^R \cos y^2\,\mathrm{d}y\right] + \left[\int_0^R \cos x^2\,\mathrm{d}x\right]\left[\int_0^R \sin y^2\,\mathrm{d}y\right].$$

Questi integrali non possono calcolarsi in modo esplicito, ma è possibile provare che il limite per $R \to +\infty$ è finito. Infatti, si consideri per esempio

$$\int_0^R \cos x^2\,\mathrm{d}x = \int_0^1 \cos x^2\,\mathrm{d}x + \int_1^{+\infty} \cos x^2\,\mathrm{d}x\,.$$

Integrando per parti, proviamo che è finito il limite[6]:

$$\lim_{R\to+\infty} \int_1^R \cos x^2\,\mathrm{d}x\,.$$

$$\int_1^R \cos x^2\,\mathrm{d}x = \int_1^R [2x\cos x^2]\frac{1}{2x}\,\mathrm{d}x = \int_1^R \frac{1}{2x}\,\mathrm{d}\sin x^2$$
$$= \frac{1}{2R}\sin R^2 - \sin 1 + \int_1^R \frac{1}{2x^2}\sin x^2\,\mathrm{d}x\,.$$

E ora,

$$\lim_{R\to+\infty} \frac{1}{2R}\sin R^2 = 0$$

mentre

$$\lim_{R\to+\infty} \int_1^R \frac{1}{2x^2}\sin x^2\,\mathrm{d}x$$

esiste finito perché $|\frac{1}{2x^2}\sin x^2| < M/x^2$ ed

$$\int_1^{+\infty} 1/x^2\,\mathrm{d}x = 1\,.$$

Si può però provare che se esiste finito

$$\lim_{R\to+\infty} \int_{D_R} |f(x,y)|\,\mathrm{d}x\ \mathrm{d}y \tag{7.8}$$

allora esiste finito

---

[6]con tecniche diverse si potrebbe anche provare che

$$\int_0^{+\infty} \sin x^2\,\mathrm{d}x = \int_0^{+\infty} \cos x^2\,\mathrm{d}x = \frac{\sqrt{\pi}}{2\sqrt{2}}\,.$$

Questi due integrali si incontrano in problemi di ottica e si chiamano *integrali di* *Fresnel*.

$$\lim_{R\to+\infty}\int_{D_R} f(x,y)\,\mathrm{d}x\ \mathrm{d}y\,.$$

Sia inoltre $\mathcal{D}_n$ una successione di domini di integrazione con questa proprietà: esiste un disco di raggio R_n contenuto in $\mathcal{D}_n$, ed $R_n\to+\infty$. Se vale (7.8) allora

$$\lim_{R\to+\infty}\int_{\mathcal{D}_R} f(x,y)\,\mathrm{d}x\ \mathrm{d}y=\lim_{R\to+\infty}\int_{D_R} f(x,y)\,\mathrm{d}x\ \mathrm{d}y\,.$$

Questo numero si chiama *integrale improprio su* $\mathbf{R}^2$ e la condizione (7.8) assicura che l'integrale improprio non dipende dal modo con cui viene calcolato, usando gli insiemi $\mathcal{D}_n$ con la proprietà suddetta[7], che $\mathcal{D}_n$ contiene un disco di raggio $R_n\to+\infty$.

Considerazioni del tutto analoghe valgono anche nel caso in cui si consideri una funzione $f(x,y)$ definita su un dominio di integrazione $\mathcal{D}$, ma tale che

$$\lim_{(x,y)\to(x_0,y_0)}|f(x,y)|=+\infty\,.$$

Sia D_ϵ un dico di raggio $\epsilon>0$ e centro (x_0,y_0) e supponiamo che $f(x,y)$ sia integrabile su $\mathcal{D}-D_\epsilon$. Supponiamo che

$$\lim_{\epsilon\to0}\int_{\mathcal{D}-D_\epsilon}|f(x,y)|\,\mathrm{d}x\ \mathrm{d}y=L<+\infty\,. \tag{7.9}$$

Allora esiste finito anche il limite seguente, che si chiama l'*integrale improprio* di $f(x,y)$ su $\mathcal{D}$:

$$\lim_{\epsilon\to0}\int_{\mathcal{D}-D_\epsilon} f(x,y)\,\mathrm{d}x\ \mathrm{d}y\,.$$

Anche in questo caso, la condizione (7.9) assicura che i dischi non hanno ruolo privilegiato nella definizione dell'integrale.

Esempio 213 Sia D il disco di centro $(0,0)$ e raggio 1 e sia

$$f(x,y)=\frac{1}{\left[\sqrt{x^2+y^2}\right]^\gamma}\,.$$

Vogliamo capire per quali valori di γ la funzione $f(x,y)$ è integrabile su D. Si deve quindi calcolare

$$\int_{D-D_\epsilon}\frac{1}{\left[\sqrt{x^2+y^2}\right]^\gamma}\,\mathrm{d}x\ \mathrm{d}y\,.$$

Passando a coordinate polari, è immediato calcolare che quest'integrale è

$$2\pi\int_\epsilon^1\rho\frac{1}{\rho^\gamma}\,\mathrm{d}\rho\,.$$

[7]invece se per esempio gli insiemi $\mathcal{D}_n$ diventano via via "più lunghi e sottili", niente può dirsi del limite degli integrali calcolati su di essi.

Il limite per $\epsilon \to 0$ esiste finito se e solo se

$$1-\gamma > -1 \qquad \text{ossia se} \qquad \gamma < 2\,.$$

Si confronti col caso delle funzioni di una variabile:

$$\int_0^1 \frac{1}{x^\gamma}\,\mathrm{d}x$$

esiste se e solo se $\gamma < 1$.

Considerazioni del tutto analoghe valgono per gli integrali tripli. In particolare, si consideri l'analogo dell'esempio 213 e si provi che

$$\int_{x^2+y^2+z^2<1} \frac{1}{\left[\sqrt{x^2+y^2+z^2}\right]^\gamma}\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z$$

esiste finito se e solo se $\gamma < 3$.

7.5.2 Appendice: Teorema dei valori intermedi e Teorema di Brower

Il Teorema dei valori intermedi per funzioni a valori reali e continue su una regione connessa è stato provato al cap. 4, si veda il Teorema 130. Mostriamo che la formula di cambiamento di variabili per gli integrali impropri permette di provare un teorema dei valori intermedi anche per funzioni da $\mathbf{R}^3$ in sé:

Teorema 214 (*dei valori intermedi*) *Sia $\phi(\mathbf{r})$ una funzione che trasforma la palla*

$$B = \{\mathbf{r} = (x,y,z) \mid ||\mathbf{r}||^2 = x^2+y^2+z^2 \le 1\}$$

in se stessa e che inoltre

- *è biunivoca;*
- *è di classe C^1*
- *verifica $\phi(\mathbf{r}) = \mathbf{r}$ nei punti della frontiera di B; ossia nei punti tali che $||\mathbf{r}|| = 1$.*

Sotto queste ipotesi, la trasformazione $\phi(\mathbf{r})$ è suriettiva.

Dim. Per assurdo, supponiamo che esista un punto $\mathbf{r_0}$ che non appartiene all'immagine di ϕ. Mostreremo in seguito che esiste una sferetta[8] B_1 di centro $\mathbf{r_0}$ nessuno dei cui punti appartiene all'immagine di ϕ. Sia $\epsilon > 0$ il raggio di B_1.

Sia $f(\mathbf{r})$ la funzione così definita:

- $f(\mathbf{r}) = 1$ se $\mathbf{r} \in B$, $||\mathbf{r} - \mathbf{r_0}|| < \epsilon/2$;
- $f(\mathbf{r}) = 0$ altrimenti.

[8]si noti che $\mathbf{r_0}$ potrebbe avere norma 1

La funzione $f(\mathbf{r})$ non è continua ma è integrabile e

$$\int_B f(x,y,z)\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z > 0\,.$$

Calcoliamo quest'integrale usando la formula di sostituzione di variabile: L'integrale è anche uguale a

$$\int_B f(\phi(x,y,z))J_\phi(x,y,z)\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\,.$$

Quest'integrale però è nullo, perché $\phi(x,y,z)$ prende valori solo nell'insieme in cui la funzione $f(x,y,z)$ è nulla.

Questa contraddizione prova che il punto $\mathbf{r_0}$ non può esistere.

Per completare la dimostrazione, mostriamo ora l'esistenza della palla B_ϵ. Proviamo che se la palla B_ϵ non esiste allora anche $\mathbf{r_0}$ appartiene all'immagine di ϕ. Supponiamo quindi che ogni palla di centro $\mathbf{r_0}$ contenga punti dell'immagine di ϕ. In tal caso, esiste una successione $\{\mathbf{r_n}\}$ di punti di B tale che $\phi(\mathbf{r_n}) \to \mathbf{r_0}$. Per il Teorema di Bolzano-Weierstrass, la successione $\{\mathbf{r_n}\}$ ammette una s.successione convergente: $\mathbf{r_{n_k}} \to \hat{\mathbf{r}}$ e

$$\lim_k \phi(\mathbf{r_{n_k}}) = \phi(\mathbf{r_0})\,.$$

Dunque, per continuità si ha

$$\phi(\hat{\mathbf{r}}) = \phi(\mathbf{r_0})\,.$$

Ciò completa la dimostrazione. ∎

Naturalmente, la dimostrazione precedente può applicarsi al caso di funzioni definite su $\mathbf{R}^2$.

Osservazione 215 E' possibile provare che il teorema precedente vale supponendo solamente che la funzione $\phi(\mathbf{r})$ sia continua e verifichi $\phi(\mathbf{r}) = \mathbf{r}$ se $||\mathbf{r}|| = 1$, senza richiedere né l'esistenza delle derivate né la biunivocità. ∎

Un corollario importante è il seguente:

Corollario 216 *Non esiste una funzione $\phi(\mathbf{r})$ continua in B e tale che*

- $\phi(\mathbf{r}) = \mathbf{r}$ se $||\mathbf{r}|| = 1$*;*
- $||\phi(\mathbf{r})|| = 1$ *per ogni* $\mathbf{r} \in B$ *.*

Dim. Infatti, una tale funzione violerebbe il teorema dei valori intermedi. ∎

Il teorema precedente ha una conseguenza importante, che va sotto il nome di *Teorema di punto fisso di Brower*.

Teorema 217 *Sia $\psi(\mathbf{r})$ una funzione continua che trasforma la palla (chiusa) B in se stessa. Esiste un punto $\mathbf{r_0} \in B$ tale che*

$$\psi(\mathbf{r_0}) = \mathbf{r_0}\,.$$

<u>Dim.</u> Accenniamo alla dimostrazione. Supponiamo che tale punto non esista. Allora, per ogni $\mathbf{r} \in B$ si ha $\psi(\mathbf{r}) \neq \mathbf{r}$. Consideriamo la semiretta di estremo $\psi(\mathbf{r})$ e che passa per $\mathbf{r}$. Questa semiretta taglia la superficie della sfera in un punto che dipende da $\mathbf{r}$. Associando ad $\mathbf{r}$ tale punto, si costruisce una funzione $\phi(\mathbf{r})$ che verifica

$$\begin{array}{ll} \|\phi(\mathbf{r})\| = 1 & \text{per ogni } \mathbf{r}\,, \\ \phi(\mathbf{r}) = \mathbf{r} & \text{se} \qquad ||\mathbf{r}|| = 1\,. \end{array}$$

Inoltre, si potrebbe provare che la funzione $\phi(\mathbf{r})$ è continua. Il Corollario 216 mostra che tale funzione non può esistere.

Si noti che la funzione $\phi(\mathbf{r})$ non è biunivoca e nemmeno di classe C^1. Però, abbiamo detto che il Teorema 214 vale con la sola ipotesi che la $\phi(\mathbf{r})$ sia continua, si veda l'osservazione 215. E quindi questa è la sola ipotesi necessaria per il Corollario 216. ∎

Capitolo 8

Integrali di curva e di superficie

Studiamo ora gli integrali definiti, invece che su intervalli o su parti di piano, su curve e su superfici. Conviene premettere alcune considerazioni sui limiti di funzioni definite su curve e, successivamente, su superfici.

In questa parte introdurremo certe notazioni e anche certi termini che non sono affatto standardizzati. Riassumiamo alcuni termini nella tabella seguente

Termine che uso io	**Altri termini**
integrale di curva di prima specie	integrale curvilineo; integrale curvilineo di prima specie.
integrale di curva di seconda specie	integrale di linea; integrale di linea di seconda specie.
integrale di superficie di prima specie	integrale superficiale.
integrale di superficie di seconda specie	integrale di superficie; integrale di flusso.

Il fatto da ricordare è questo: gli integrali di prima specie integrano funzioni (a valori reali) definite su curve o su superfici; gli integrali di seconda specie integrano campi vettoriali definiti su curve o su superfici.

N.B. Per ragioni tipografiche, alcune figure di questo capitolo si trovano alla fine del capitolo stesso.

8.1 Funzioni definite su curve: la densità

Funzioni definite su curve, anche a valori vettoriali, si sono già incontrate: abbiamo incontrato il campo vettoriale che ad ogni punto di una curva associa la tangente oppure la normale alla curva nel punto stesso. Queste funzioni possono intendersi come funzioni definite, per esempio, su $\mathbf{R}^3$, con dominio il sostegno della curva. La definizione dei limiti per tali funzioni presenta però una particolarità che va evidenziata.

Sia γ una curva semplice e regolare, parametrizzata da

$$t \to \mathbf{r}(t)\,, \qquad t \in [a,b].$$

La curva è semplice. Dunque, ogni coppia di punti $\mathbf{r_0}$ e $\mathbf{r_1}$ di γ individua un arco

$$\mathbf{r}(t)\,, \qquad s \in [t_0, t_1]\,,$$

con t_0 ed t_1 gli unici valori del parametro che verificano

$$\mathbf{r_0} = \mathbf{r}(t_0)\,, \qquad \mathbf{r_1} = \mathbf{r}(t_1)\,.$$

Introduciamo la notazione $\gamma_{\mathbf{r_0},\mathbf{r_1}}$ per indicare quest'arco e la notazione $L_{\gamma(\mathbf{r_0},\mathbf{r_1})}$ per indicarne la lunghezza (si veda il paragrafo (6.1.2):

$$L_{\gamma(\mathbf{r_0},\mathbf{r_1})} = \int_{t_0}^{t_1} |\mathbf{r}'(\nu)| \, \mathrm{d}\nu\,.$$

Sia $f(\mathbf{r})$ una funzione definita nei punti del sostegno di γ. Chiamiamo *densità* della funzione f nel punto $\mathbf{r_0} = \mathbf{r}(t_0) \in \gamma$ il limite seguente:

$$\rho_f(\mathbf{r_0}) = \lim_{h\to 0} \frac{f(\mathbf{r}(t_0+h)) - f(\mathbf{r}(t_0))}{L_{\gamma(\mathbf{r}(t_0),\mathbf{r}(t_0+h))}}\,. \tag{8.1}$$

Notare che il rapporto incrementale ha per denominatore non lo scarto h del parametro, ma la lunghezza dell'arco che congiunge $\mathbf{r}(t_0)$ con $\mathbf{r}(t_0+h)$. Naturalmente, il denominatore è lo scarto del parametro se il parametro è s, il parametro d'arco.

Se l'arco rappresenta un "filo" di materiale non omogeneo e se $f(\mathbf{r})$ rappresenta la massa totale tra l'estremo $\mathbf{r}(a)$ e il punto $\mathbf{r}$ della curva, allora il limite precedente, se esiste, rappresenta la usuale densità di massa del "filo". Però, la funzione $f(\mathbf{r})$ non è necessariamente positiva, e potrebbe essere una componente di un campo vettoriale, caso che si è già incontrato definendo la normale a γ in $\mathbf{r}_0$. Infatti, la definizione di limite (8.1) è quella usata nella definizione del versore normale.

8.2 Gli integrali di curva

Integrali di funzioni definite su curve possono definirsi in vari modi. Per dare definizioni significative, è necessario avere come guida degli esempi tratto dalla fisica. Prendiamo come

guida il problema di calcolare la massa totale di un "filo", nota la sua densità e il problema di calcolare il lavoro di una campo di forze su un punto che si muove lungo una curva. Il primo esempio conduce alla definizione di integrale di curva di prima specie mentre il secondo esempio conduce alla definizione di integrale di curva di seconda specie.

8.2.1 Integrali di curva di prima specie

Consideriamo l'esempio seguente:

Esempio 218 supponiamo che $t \to \mathbf{r}(t)$, $t \in [a, b]$ parametrizzi un arco realizzato con un materiale non omogeneo, di densità $\rho(\mathbf{r})$. Ricordando la definizione di densità, la massa dell'arco si approssima dividendolo in segmenti di lunghezza δ, molto piccola, e sommando la massa dei singoli pezzetti. Il modo migliore di fare ciò, anche se non necessariamente il più semplice dal punto di vista del calcolo, consiste nel ricondursi alla parametrizzazione canonica, rappresentando l'arco come $s \to \mathbf{r}(s)$, $s \in [0, L]$; dividere $[0, L]$ con N punti s_i, $s_i = iL/N$ e quindi costruire

$$\sum_{i=0}^{N-1} \rho(\mathbf{r}(s_i))[s_{i+1} - s_i] = \sum_{i=0}^{N-1} \rho(\mathbf{r}(s_i))L/N\,.$$

Studiare quindi il comportamento di queste somme per $N \to +\infty$, ossia quando la finezza della suddivisione dell'arco tende a zero. Questa è niente altro che la costruzione dell'integrale della funzione $s \to \rho(\mathbf{r}(s))$ sull'intervallo $[0, L]$. ∎

Sia ora $f(\mathbf{r})$ una generica funzione, per semplicità continua. L'esempio precedente suggerisce di definire

$$\int_\gamma f(\mathbf{r})\,\mathrm{d}s$$

come segue: prima rappresentiamo l'arco in forma canonica, ossia mediante il parametro d'arco e quindi definiamo

$$\int_\gamma f(\mathbf{r})\,\mathrm{d}s =_{\text{def}} \int_0^L f(\mathbf{r}(s))\,\mathrm{d}s\,. \tag{8.2}$$

Questa definizione richiede che l'arco γ sia regolare, ed è suggerita dal significato fisico che vogliamo attribuire all'integrale. Per il calcolo pratico conviene però evitare di rappresentare in forma canonica l'arco. Conviene di più lavorare con la parametrizzazione $\mathbf{r}(t)$, $t \in [a, b]$, inizialmente assegnata. Notiamo che si passa dalla parametrizzazione $\mathbf{r} = \mathbf{r}(t)$ alla parametrizzazione canonica per mezzo del cambiamento di variabile $t = t(s)$. Dunque l'integrale a destra di (8.2) è, in realtà,

$$\int_0^L f(\mathbf{r}(t(s)))\,\mathrm{d}s\,.$$

Ossia, in (8.2), solo per semplicità di notazioni, abbiamo sostituito il simbolo $\mathbf{r}(s)$ alla notazione più completa $\mathbf{r}(t(s))$.

La funzione $s = s(t)$, ossia la funzione inversa della funzione $t(s)$, è derivabile, con derivata

$$s'(t) = |\mathbf{r}'(t)|\,.$$

La sostituzione di variabile $s = s(t)$ nell'integrale a destra di (8.2) mostra che vale:

$$\int_\gamma f(\mathbf{r})\,\mathrm{d}s = \int_a^b f(\mathbf{r}(t))\,|\mathbf{r}'(t)|\,\mathrm{d}t\,. \tag{8.3}$$

Se l'arco è in $\mathbf{R}^3$, quest'integrale è

$$\int_a^b f(\mathbf{r}(t))\sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}\,\mathrm{d}t\,.$$

Se l'arco è cartesiano, parametrizzata da

$$y = y(x)\,, \quad z = z(x)\,, \qquad x \in [a,b]$$

allora

$$\int_\gamma f(\mathbf{r})\,\mathrm{d}s = \int_a^b f(x, y(x), z(x))\sqrt{1 + [y'(x)]^2 + [z'(x)]^2}\,\mathrm{d}x\,.$$

L'integrale appena definito si chiama *integrale di curva* *di prima specie*. La definizione stessa mostra che esso non dipende dalla parametrizzazione scelta per rappresentare la curva. Più ancora, se si cambia la variabile t mediante la trasformazione $t = t(\tau) = b + a - \tau$, $\tau \in [a,b]$, il valore dell'integrale non cambia, ossia:

Teorema 219 *l'integrale di curva di prima specie non cambia né cambiando la parametrizzazione né cambiando il verso di percorrenza dell'arco.*

Il fatto che l'integrale di curva di prima specie non dipenda dalla parametrizzazione mostra che, per ogni fissata funzione $f(\tau)$, esso è una proprietà geometrica della curva.

Il fatto che l'integrale di curva di prima specie non cambi valore cambiando il verso di percorrenza si scrive come segue:

$$\int_{-\gamma} f(\mathbf{r})\,\mathrm{d}s = \int_\gamma f(\mathbf{r})\,\mathrm{d}\mathbf{s}$$

Notiamo infine la disuguaglianza seguente:

Teorema 220 *Sia $f(\mathbf{r})$ una funzione continua e sia $|f(\mathbf{r})| < M$ in ogni punto $\mathbf{r}$ del sostegno di γ. Allora,*

$$\left|\int_\gamma f(\mathbf{r})\,ds\right| \le M L_\gamma\,.$$

In particolare, si consideri un arco γ parametrizzato da

$$\gamma\ :\ t \to \mathbf{r}(t)\,, \qquad t \in [a,b]$$

ed una sua parte γ_ϵ parametrizzata da

$$\gamma_\epsilon\ :\ t \to \mathbf{r}(t)\,, \qquad t \in [a,\epsilon]\,.$$

Vale

$$\lim_{\epsilon \to a+} \int_{\gamma_\epsilon} f(\mathbf{r})\,\mathrm{d}s = 0\,. \tag{8.4}$$

Notiamo infine che se si divide un arco γ γ_1 e γ_2,

$$\gamma\ : t \to \mathbf{r}(t)\,, \quad t \in [a,c]$$
$$\gamma_1\ : t \to \mathbf{r}(t)\,, \quad t \in [a,b]\,; \qquad \gamma_2\ : t \to \mathbf{r}(t)\,, \quad t \in [b,c]\,;$$

allora vale

$$\int_\gamma f(\mathbf{r})\,\mathrm{d}s = \int_{\gamma_1} f(\mathbf{r})\,\mathrm{d}s + \int_{\gamma_2} f(\mathbf{r})\,\mathrm{d}s\,. \tag{8.5}$$

Osservazione 221 La definizione di integrale di curva di prima specie è stata data assumendo che l'arco sia regolare. La (8.5) mostra come definire l'integrale nel caso di un arco regolare a tratti: se l'arco γ si "decompone" in (per esempio) due sottoarchi γ_1 e γ_2 regolari, si sceglie la (8.5) come definizione di integrale su γ. ▮

8.2.2 Integrali di curva di seconda specie

La definizione che ora andiamo a dare generalizza quella che si usa in fisica per il calcolo di un lavoro.

Esempio 222 Supponiamo che in ogni punto $\mathbf{r}$ dello spazio agisca una forza $\mathbf{F}(\mathbf{r}) = \mathbf{F}(x,y,z) = f(x,y,z)\mathbf{i} + g(x,y,z)\mathbf{j} + h(x,y,z)\mathbf{k}$. Un punto materiale di massa m sia vincolato a descrivere un arco $\mathbf{r}(t)$, $t \in [a,b]$. Il parametro t rappresenta ora il tempo. Si vuol valutare il lavoro compiuto dalla forza.

Dividendo ancora il percorso del punto in tanti "piccoli pezzetti", interessa ora valutare il prodotto scalare della forza agente su ciascun "pezzetto" per lo spostamento del punto. Lo spostamento è un vettore, a differenza della lunghezza percorsa che è un numero.

Quando t varia da t_i a t_{i+1} lo spostamento è circa $\mathbf{r}'(t_i)(t_{i+1} - t_i)$ e quindi il lavoro compiuto è circa

$$[\,\mathbf{F}(\mathbf{r}(t_i)) \cdot \mathbf{r}'(t_i)\,](t_{i+1} - t_i)\,.$$

La somma di tutti questi "lavori elementari" è

$$\sum_{i=0}^{N-1} [\mathbf{F}(\mathbf{r}(t_i)) \cdot \mathbf{r}'(t_i)](t_{i+1} - t_i)$$

e, per calcolare il lavoro della forza bisogna studiare il comportamento di queste somme, quando la finezza della suddivisione dell'intervallo $[a,b]$ tende a zero. Si sa che in questo modo si arriva a calcolare l'integrale su $[a,b]$ della funzione

$$\mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\,. \quad ▮$$

Seguendo il suggerimento dell'esempio precedente definiamo l'*integrale di curva* *di seconda specie* come segue: si assegna un campo vettoriale $\mathbf{V}(\mathbf{r})$ e un arco γ, parametrizzato da $\mathbf{r}(t)$, $t \in [a,b]$. Si definisce

$$\int_\gamma \mathbf{V}(\mathbf{r}) \cdot \,\mathrm{d}\mathbf{r} =_{\mathrm{def}} \int_a^b \mathbf{V}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\,\mathrm{d}t\,. \tag{8.6}$$

Nel caso particolare $n = 3$ e $\mathbf{V}(x, y, z) = f(x, y, z)\mathbf{i} + g(x, y, z)\mathbf{j} + h(x, y, z)\mathbf{k}$ si trova:

$$\int_\gamma \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \int_a^b [f(x(t), y(t), z(t))x'(t) \\ +g(x(t), y(t), z(t))y'(t) + h(x(t), y(t), z(t))z'(t)] \, \mathrm{d}t \, . \tag{8.7}$$

Osservazione 223 Una curva semplice e chiusa si chiama anche un *circuito* e l'integrale di un campo vettoriale $\mathbf{V}(\mathbf{r})$ lungo una curva chiusa si chiama anche la *circuitazione* del campo vettoriale lungo γ. Quando la curva γ è semplice e chiusa, l'integrale su γ si indica anche col simbolo

$$\oint_\gamma \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} \, . \quad \blacksquare$$

E' immediato verificare, usando la formula di cambiamento di variabili:

Teorema 224 *Il valore dell'integrale di curva di seconda specie non muta cambiando parametrizzazione; cambia di segno cambiando verso di percorrenza.*

La prima affermazione mostra che, per ogni fissato campo vettoriale $\mathbf{V}(\mathbf{R})$, l'integrale di curva di seconda specie è una proprietà geometrica della curva. La seconda affermazione si esprime scrivendo

$$\int_{-\gamma} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = -\int_\gamma \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} \, .$$

E' ovvio inoltre che se si divide un arco γ in due archi γ_1 e γ_2,

$$\gamma \, : t \to \mathbf{r}(t) \, , \quad t \in [a, c]$$
$$\gamma_1 \, : t \to \mathbf{r}(t) \, , \quad t \in [a, b] \, ; \qquad \gamma_2 \, : t \to \mathbf{r}(t) \, , \quad t \in [b, c] \, ;$$

allora vale

$$\int_\gamma \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \int_{\gamma_1} V(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} + \int_{\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} \, . \tag{8.8}$$

Esattamente come nell'osservazione 221, la (8.8), può usarsi per estendere la definizione di integrale di curva di seconda specie ad archi regolari a tratti.

L'arco γ ottenuto percorrendo prima γ_1 e poi γ_2 si indica col simbolo $\gamma_1 + \gamma_2$. Questo simbolo non indica soltanto l'unione insiemistica dei sostegni perché per il calcolo dell'integrale bisogna anche specificare il verso di percorrenza. Con questa notazione, la (8.8) si scrive anche

$$\int_{\gamma_1+\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \int_{\gamma_1} V(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} + \int_{\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} \, .$$

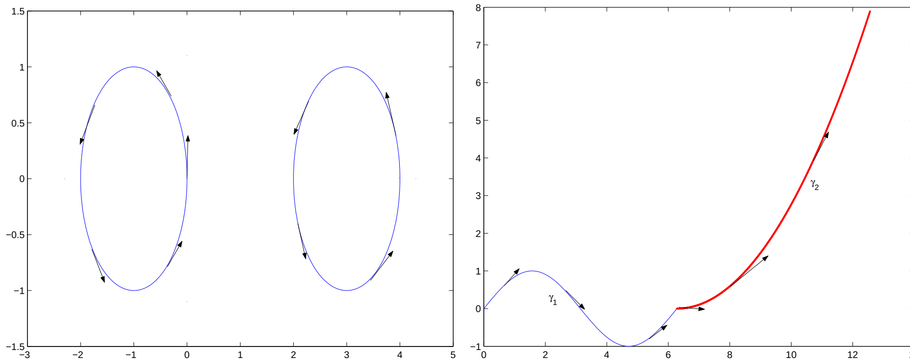
Quest'osservazione suggerisce la notazione seguente: siano γ_1 e γ_2 due archi, non necessariamente "sottoarchi" del medesimo. Col simbolo $\gamma_1 + \gamma_2$ intendiamo di percorrere prima l'arco γ_1, preso col suo verso di percorrenza, e quindi l'arco γ_2 preso col suo verso di percorrenza.

Col simbolo $\gamma_1 - \gamma_2$ si intende di percorrere prima l'arco γ_1, col suo verso di percorrenza, e poi l'arco γ_2 col verso di percorrenza negativo. Definiamo quindi

$$\begin{aligned}
&\int_{\gamma_1+\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \int_{\gamma_1} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} + \int_{\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r}\,, \\
&\int_{\gamma_1-\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \int_{\gamma_1} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} + \int_{-\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \\
&= \int_{\gamma_1} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} - \int_{\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r}\,.
\end{aligned}$$

Consideriamo ora la figura 8.1. La figura a sinistra rappresenta due archi, γ_1 e γ_2 tali che né $\gamma_1 + \gamma_2$ né $\gamma_1 - \gamma_2$ rappresentano un arco. La figura a destra rappresenta due archi, ciascuno col proprio verso di percorrenza, tali che $\gamma_1 + \gamma_2$ rappresenta un arco mentre $\gamma_1 - \gamma_2$ non rappresenta un arco.

Figura 8.1: "Operazioni" sugli archi



La figura 8.2 rappresenta a sinistra l'arco γ_1

$$\gamma_1 \;:\; x = (1-t)^2\,, \quad y = (1-t)^3\,, \qquad t \in [0,1] \tag{8.9}$$

e, nel medesimo piano, l'arco γ_2:

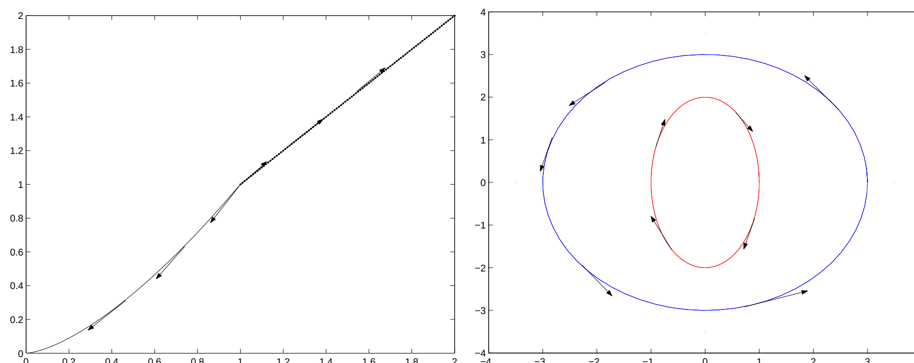
$$\gamma_2 \;:\; x = t\,, \quad y = t\,, \qquad t \in [1,2] \tag{8.10}$$

ciascuno con indicato il verso di percorrenza. E' chiaro che $\gamma_1 + \gamma_2$ non è un arco, mentre $\gamma_1 - \gamma_2$ lo è. La figura a destra rappresenta un caso tipico, che useremo più volte. Si noti che la curva "esterna" e la curva "interna" sono percorse in verso opposto.

Per gli integrali di curva di seconda specie può darsi un risultato analogo alla (8.4). Usando le medesime notazioni, si ha:

$$\lim_{\epsilon \to a+} \int_{\gamma_\epsilon} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = 0\,, \tag{8.11}$$

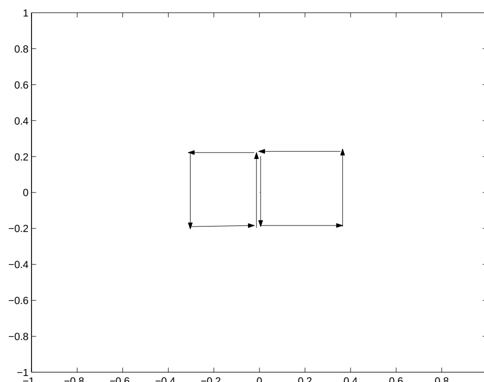
Figura 8.2: Gli archi (8.9) e (8.10) a sinistra. A destra una regione delimitata da due archi



fatto che si prova facilmente introducendo la parametrizzazione dell'arco[1]. Ricordando però che l'integrale di curva di seconda specie cambia segno cambiando il verso di percorrenza, si possono enunciare due risultati che non hanno analogo per gli integrali di curva di prima specie.

Supponiamo che un arco ripassi due volte su un arco γ_1, percorrendolo in versi opposti. Allora, nel calcolo dell'integrale γ_1 non dà contributo. In particolare, si consideri la figura 8.3. Nella figura, i due lati affiancati vanno pensati sovrapposti e sono il sostegno dell'arco γ_1. Sono stati disegnati soltanto affiancati per chiarezza.

Figura 8.3: Archi che "si elidono"



[1]come diremo al paragrafo 8.2.3, questa proprietà immediatamente discende dalla (8.4).

Indichiamo con γ_s la curva il cui sostegno è il quadrato di sinistra e con γ_d quella il cui sostegno è il quadrato di destra. Nella somma

$$\int_{\gamma_s} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} + \int_{\gamma_d} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r}$$

il contributo di γ_1 si elide e si trova

$$\int_{\gamma_s+\gamma_d} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \int_{\gamma_s} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} + \int_{\gamma_d} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \int_{\gamma} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} \tag{8.12}$$

ove γ è la curva il cui sostegno è il rettangolo che si ottiene sopprimendo il lato comune ai due quadrati, col verso di percorrenza indicato.

Naturalmente nelle considerazioni precedenti il fatto che le curve siano ottenute per mezzo di segmenti rettilinei non ha alcuna importanza.

Una seconda osservazione, semplice ma importante, è la seguente: siano γ_0 e γ_τ (τ è un parametro) due segmenti

$$\gamma \,:\mathbf{r} = \mathbf{r_0} + t\mathbf{u_0}\,, \quad t \in [0, a] \qquad \gamma_\tau \,:\mathbf{r} = \mathbf{r} + t\mathbf{u}\,, \quad t \in [0, b]\,.$$

Supponiamo che

$$||\mathbf{u_0}|| = 1\,, \qquad ||\mathbf{u}|| = 1\,.$$

e consideriamo i due integrali

$$\int_{\gamma} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r}\,, \qquad \int_{\gamma_\tau} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r}\,.$$

Supponiamo che il campo vettoriale sia continuo e che "il segmento γ_τ tenda a sovrapporsi a γ_0". Questo vuol dire che il secondo segmento dipende da un parametro, diciamo $\tau \in [0, 1]$, ossia che

$$\mathbf{r} = \mathbf{r}(\tau)\,, \quad \mathbf{u} = \mathbf{u}(\tau)\,, \quad b = b(\tau)\,.$$

Assumiamo che queste funzioni dipendano da τ in modo continuo e che sia

$$\mathbf{r}(0) = \mathbf{r}_0\,, \quad \mathbf{u}(0) = \mathbf{u}_0\,, \quad b(0) = a\,.$$

In tal caso:

Teorema 225 *Si ha:*

$$\lim_{\tau \to 0} \int_{\gamma_\tau} \mathbf{V}(\mathbf{r}) \cdot \, d\mathbf{r} = \int_{\gamma_0} \mathbf{V}(\mathbf{r}) \cdot \, d\mathbf{r}\,.$$

La dimostrazione è ovvia. Infatti,

$$\Phi(t, \tau) = [\mathbf{V}(\mathbf{r}(\tau) + t\mathbf{u}(\tau))] \cdot \mathbf{u}(\tau)$$

è una funzione continua delle due variabili t e τ e inoltre

$$\Phi(t, 0) = \mathbf{V}(\mathbf{r_0} + t\mathbf{u_0}) \cdot \mathbf{u_0}\,.$$

Dunque,

$$\lim_{\tau \to 0} \int_{\gamma_\tau} \mathbf{V}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \lim_{\tau \to 0} \int_0^{b(\tau)} \Phi(t, \tau) \, \mathrm{d}t = \int_0^a \Phi(t, 0) \, \mathrm{d}1 = \int_{\gamma} \mathbf{V}(\mathbf{r}) \cdot \mathbf{r}\,,$$

si veda il paragrafo 4.2.4.

In particolare,

$$\lim_{\tau \to 0} \int_{-\gamma_\tau} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = -\int_{\gamma_0} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r}\,.$$

Dunque,

Corollario 226 *Nelle ipotesi dette sopra si ha*

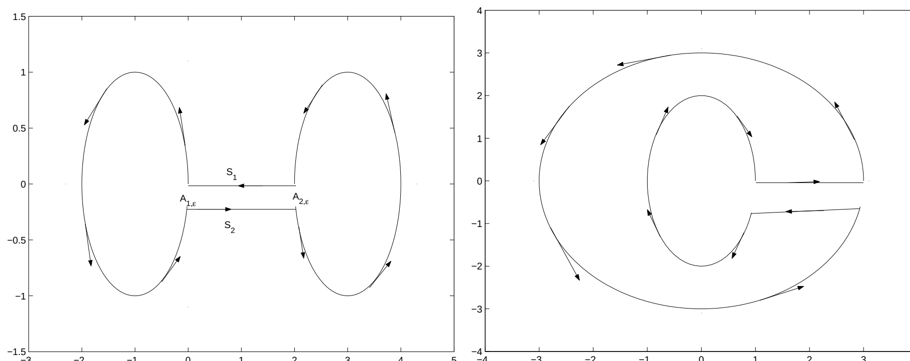
$$\lim_{\tau \to 0} \left[\int_{-\gamma_\tau} \mathbf{V}(\mathbf{r}) \cdot\, d\mathbf{r} + \int_{\gamma_0} \mathbf{V}(\mathbf{r}) \cdot\, d\mathbf{r} \right] = 0\,.$$

Ossia, Il contributo dei due integrali tende ad elidersi. Questo è il corollario che useremo più avanti. Ora, usiamo sia la (8.11) che il Corollario 226 per giustificare l'uso della notazione

$$\int_{\gamma_1 + \gamma_2}$$

anche nel caso in cui $\gamma_1 + \gamma_2$ non è un arco. Torniamo a considerare la figure 8.1 a sinistra. Si confronti questa figura con la figura 8.4, a sinistra. La figura di sinistra rappresenta

Figura 8.4: Integrali di curva di seconda specie ed archi che "si elidono"



una curva semplice e chiusa, chiamiamola γ_ϵ, dipendente dal parametro ϵ, lunghezza degli archi (indicati con $A_{1,\epsilon}$, $A_{2,\epsilon}$) "rimossi" dalle ellissi di destra e di sinistra, che chiamiamo rispettivamente γ_1 e γ_2. Dunque,

$$\begin{aligned} &\int_{\gamma_\epsilon} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} - \int_{\gamma_1+\gamma_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \\ &= \int_{S_1} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} + \int_{S_2} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \\ &- \int_{A_{1,\epsilon}} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} - \int_{A_{2,\epsilon}} \mathbf{V}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r}\,. \end{aligned}$$

Facendo tendere ϵ a zero, l'integrale sugli archi $A_{1,\epsilon}$, $A_{2,\epsilon}$ tende a zero per la (8.11), mentre la somma degli integrali sui due segmenti tende a zero per il Corollario 226. Dunque,

$$\lim_{\epsilon\to 0}\int_{\gamma_\epsilon} \mathbf{V}(\mathbf{r})\cdot\, \mathrm{d}\mathbf{r} = \int_{\gamma_1+\gamma_2} \mathbf{V}(\mathbf{r})\cdot\, \mathrm{d}\mathbf{r}\,.$$

Ciò spega perché è necessario considerare espressioni del tipo

$$\int_{\gamma_1+\gamma_2}$$

anche nel caso in cui $\gamma_1+\gamma_2$ non è un arco.

La figura 8.4, a destra, mostra una spiegazione analoga nel caso illustrato nella figura 8.1, a destra.

Osservazione 227 Si noti che in ambedue gli esempi, gli archi i cui integrali approssimano quello su $\gamma_1+\gamma_2$ sono semplici. Fissato il verso di percorrenza su γ_1, quello su γ_2 viene automaticamente determinato se si vuol avere un'approssimazione con un arco semplice. In particolare, nel caso delle due curve di Jordan, una nella regione interna dell'altra, usa prendere la curva più esterna orientata positivamente e quindi quella più interna orientata negativamente. Più avanti vedremo la ragione di tale scelta. ∎

Infine, supponiamo di avere due curve γ_1 e γ_2, non necessariamente semplici:

$$\gamma_1\ :\ t\to\mathbf{r}(t)\quad t\in[a,b]\,;\qquad \gamma_2\ :\ \tau\to\rho(\tau)\quad \tau\in[\alpha,\beta]\,.$$

Diciamo che le due curve hanno *gli stessi* *estremi* quando

$$\mathbf{r}(a)=\rho(\alpha)\,,\qquad \mathbf{r}(b)=\rho(\beta)\,.$$

In questa definizione si è tenuto conto del verso di percorrenza delle curve: i due "primi estremi" devono coincidere ed i due "secondi estremi" devono coincidere. In tal caso:

Lemma 228 *Nelle ipotesi dette sopra, $\gamma_1-\gamma_2$ è una curva chiusa e*

$$\int_{\gamma_1-\gamma_2} \mathbf{V}(\mathbf{r})\cdot\, d\mathbf{r} = \int_{\gamma_1} \mathbf{V}(\mathbf{r})\cdot\, d\mathbf{r} - \int_{\gamma_2} \mathbf{V}(\mathbf{r})\cdot\, d\mathbf{r}\,.$$

8.2.3 Integrali di curva di prima e di seconda specie

E' interessante confrontare le definizioni degli integrali di curva di prima e seconda specie nel caso particolare in cui $\mathbf{V}(\mathbf{r})=f(\mathbf{r})\mathbf{i}$. In questo caso

$$\int_\gamma \mathbf{V}(\mathbf{r})\cdot\, \mathrm{d}\mathbf{r} = \int_a^b f(x(t),y(t),z(t))x'(t)\,\mathrm{d}t$$

mentre

$$\int_\gamma f\,\mathrm{d}s = \int_a^b f(x(t),y(t),z(t))\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}\,\mathrm{d}t\,.$$

Questi due integrali non si riconducono l'uno all'altro nemmeno nel caso in cui $y(t)$ e $z(t)$ sono identicamente zero. Infatti, in tal caso essi vengono ad essere, rispettivamente,

$$\int_a^b f(x(t),y(t),z(t))\,x'(t)\,\mathrm{d}t\,,\qquad \int_a^b f(x(t),y(t),z(t))|x'(t)|\,\mathrm{d}t\,.$$

Ciò nonostante, esiste una relazione tra gli integrali di prima e seconda specie, che ora mostriamo.

Consideriamo l'arco

$$\gamma:\quad x(t)\mathbf{i}+y(t)\mathbf{j}+z(t)\mathbf{k}\,,\qquad t\in[a,b]$$

Ricordiamo che con $\mathbf{t}(\mathbf{r})$ si è indicato il versore tangente alla curva nel punto $\mathbf{r}$:

$$\mathbf{t}(\mathbf{r}(t))=\frac{\mathbf{t}'(t)}{|\mathbf{r}'(t)|}=\frac{x'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}\mathbf{i}$$
$$+\frac{y'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}\mathbf{j}+\frac{z'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}\mathbf{k}$$

Sia

$$\mathbf{V}(\mathbf{r})=V_1(\mathbf{r})\mathbf{i}+V_2(\mathbf{r})\mathbf{j}+V(\mathbf{r})\mathbf{k}$$

un campo vettoriale definito sulla curva γ. Introduciamo la funzione

$$f(t)=\mathbf{V}(\mathbf{r}(t))\cdot\left\{\frac{x'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}\mathbf{i}\right.$$
$$\left.+\frac{y'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}\mathbf{j}+\frac{z'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}\mathbf{k}\right\}.$$

Si ha

$$\int_\gamma \mathbf{V}(\mathbf{r})\cdot\,\mathrm{d}\mathbf{r}=\int_\gamma f(\mathbf{r})\,\mathrm{d}s\,. \tag{8.13}$$

Talvolta questa formula si trova scritta in modo diverso. Notiamo che

$$n_1=\frac{x'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}$$
$$n_2=\frac{y'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}$$
$$n_3=\frac{z'(t)}{\sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}}$$

sono i coseni direttori della tangente $\mathbf{t}(\mathbf{r}(t))$. Per questo la formula (8.13) si trova anche scritta come segue:

$$\int_\gamma \mathbf{V}(\mathbf{r})\cdot\,\mathrm{d}\mathbf{r}=\int_\gamma [V_1n_1+V_2n_2+V_3n_3]\,\mathrm{d}s\,.$$

Notiamo infine che la (8.11) si può derivare dalla (8.4), usando la (8.13).

8.2.4 Integrali di curva di seconda specie e forme differenziali

L'integrale di curva di seconda specie si calcola quando è dato un campo vettoriale $\mathbf{V}(\mathbf{r})$. Dunque, possiamo vedere il campo vettoriale come operante sulla curva γ: il campo vettoriale $\mathbf{V}(\mathbf{r})$ associa un numero alla curva γ. Vedremo che altri campi vettoriali si usano per associare numeri alle superfici. Dunque conviene distinguere i due casi, introducendo termini e notazioni diverse. Nel caso che stiamo considerando, il campo vettoriale agisce su una curva. Invece di indicarlo col simbolo $\mathbf{V}(x,y,z) = f(x,y,z)\mathbf{i} + g(x,y,z)\mathbf{j} + h(x,y,z)\mathbf{k}$, indichiamolo col simbolo[2]

$$f(x,y,z)\,\mathrm{d}x + g(x,y,z)\,\mathrm{d}y + h(x,y,z)\,\mathrm{d}z\,. \tag{8.14}$$

Non vogliamo dare nessun significato particolare ai simboli $\mathrm{d}x$, $\mathrm{d}y$ e $\mathrm{d}z$. Essi, come si è notato, sono suggeriti dalla formula (8.7). Ricordiamo però che nel caso dell'integrale su un intervallo, si conserva la notazione "$\mathrm{d}x$" perché questa aiuta a ricordare certe formule. Come vedremo, qualcosa di analogo accade anche in questo caso.

La (8.14) si chiama *1–forma differenziale* (il numero 1 ricorda che si agisce su una curva, un oggetto che, intuitivamente[3], ha dimensione 1, come un segmento o un filo). L'integrale di curva di seconda specie si chiama anche *l'integrale di curva della 1–forma differenziale* e si indica anche col simbolo

$$\int_\gamma f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z$$

(sottintendendo la dipendenza di f, g e h da x, y e z).

Una regola mnemonica per ottenere la formula (8.7), per esempio nel caso $n = 3$ è la seguente: si ricordi che le funzioni f, g ed h dipendono da (x,y,z). Allora, si sostituisca x con $x(t)$, y con $y(t)$, z con $z(t)$ ovunque queste lettere compaiono; e si interpreti "d" come segno di derivata; e quindi $\mathrm{d}x$ si sostituisce con $x'(t)\,\mathrm{d}t$, $\mathrm{d}y$ con $y'(t)\,\mathrm{d}t$ e $\mathrm{d}z$ con $z'(t)\,\mathrm{d}t$. Infine, si integri da a fino a b, ottenendo la (8.7).

Questa è una prima buona ragione per usare la notazione delle forme differenziali. Più avanti ne vedremo altre.

Una forma differenziale si indica spesso con una lettera greca minuscola tratta dalla fine dell'alfabeto, come ω,

$$\omega = f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z\,.$$

Con tale notazione l'integrale della forma differenziale si indica

$$\int_\gamma \omega\,.$$

Infine, diciamo che la forma differenziale

$$\omega = f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z$$

[2] suggerito dalla formula (8.7).

[3] e solo sotto condizioni di regolarità: esistono curve dalla parametrizzazione continua ma non derivabile, che hanno per sostegno un quadrato. Il prototipo di tali curve si chiama "curva di Peano".

è di classe C^1 quando sono di classe C^1 i suoi *coefficienti*, ossia le funzioni $f(x, y, z)$, $g(x, y, z)$, e $h(x, y, z)$.

Osservazione 229 Si noti che il simbolo dx è una forma differenziale: quella di coefficienti $f = 1$ e $g = h = 0$. Questa forma differenziale si chiama anche il *differenziale* della variabile x. Questo termine "differenziale" ha un significato diverso da quello introdotto al Cap. 4. Esistono relazioni tra questi due diversi concetti, che però non illustriamo. ∎

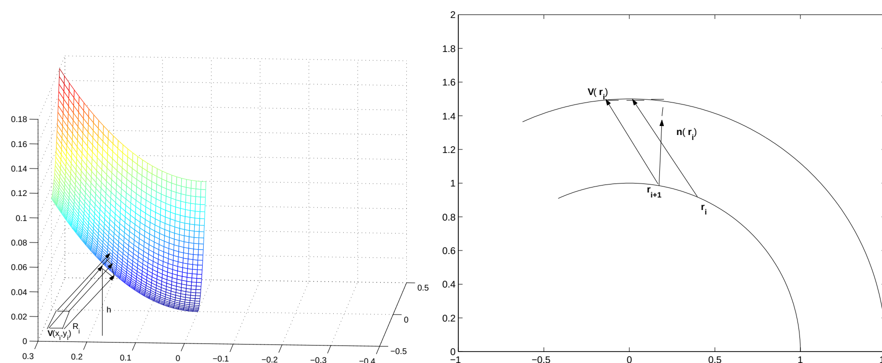
8.2.5 Il flusso

Sia

$$\mathbf{V}(x, y) = f(x, y)\mathbf{i} + g(x, y)\mathbf{j} + h(x, y)\mathbf{k}$$

un campo vettoriale che ad ogni punto del piano $z = 0$ associa un vettore di $\mathbf{R}^3$. Sia $\mathcal{D}$ un dominio di integrazione e supponiamo che $\mathbf{V}(x, y)$ sia continuo sulla chiusura di $\mathcal{D}$. Possiamo pensare che il piano $z = 0$ sia immerso in un fluido che scorre in modo che quando una particella si trova nel punto $(x, y, 0)$ ivi abbia la velocità $\mathbf{V}(x, y)$ e mantenga tale velocità dopo che ha lasciato il piano $z = 0$. Il *flusso attraverso $\mathcal{D}$* è la quantità di fluido che traversa $\mathcal{D}$ nell'unità di tempo. Per calcolarla[4] procediamo in questo modo. Approssimiamo $\mathcal{D}$ mediante un numero finito di rettangoli R_i, come nella definizione dell'integrale doppio. Una particella che si trova in un punto $(x, y, 0) \in R_i$ descrive nell'unità di tempo il segmento rettilineo che congiunge $(x, y, 0)$ con $V(x, y)$. Infatti, abbiamo detto che la velocità non cambia dopo che la particella ha lasciato il piano $z = 0$. Dunque, nell'unità di tempo le particelle che escono dal rettangolo R_i riempiono una specie di parallelepipedo, con "faccia superiore" non piana, di base R_i. L'altezza varia da punto a punto, come in figura 8.5, a sinistra. Se il rettangolo è "piccolo" l'altezza sarà circa uguale a $\mathbf{k} \cdot \mathbf{V}(x_i, y_i)$ ove (x_i, y_i)

Figura 8.5: Il flusso



è per esempio il vertice in basso a sinistra del rettangolo R_i. Dunque il volume di questo

[4]più correttamente dovremmmo dire "per definirla".

solido è approssimato da

$$h(\text{Area di } R_i) = [\mathbf{k} \cdot \mathbf{V}(x_i, y_i)](\text{Area di } R_i).$$

Si noti che l'area di R_i è positiva mentre l'altezza può anche essere negativa; ossia, il volume si prende "con segno" perché non è indifferente che il fluido scorra verso l'alto o verso il basso.

Sommiamo ora i contributi di tutti i singoli rettangoli. Si trova una somma integrale di quelle che definiscono l'integrale

$$\int_{\mathcal{D}} \mathbf{V}(x, y) \cdot \mathbf{k} \, \mathrm{d}x \, \mathrm{d}y \, .$$

Scegliamo quindi questa formula come definizione di flusso *attraverso il dominio di integrazione $\mathcal{D}$.*

Più avanti vedremo la definizione di flusso attraverso una superficie non piana. Ora, definiamo il flusso di un campo piano attraverso una curva.

Supponiamo che uno strato liquido riempia il piano (x, y) e che il campo vettoriale

$$\mathbf{V}(x, y) = f(x, y)\mathbf{i} + g(x, y)\mathbf{j}$$

rappresenti la velocità con cui una particella traversa la posizione (x, y). Sia γ un arco regolare, parametrizzata da

$$\gamma : \ x = x(t) \, , \quad y = y(t) \, , \qquad t \in [a, b] \, .$$

Si chiama *flusso* del campo $\mathbf{V}$ attraverso la curva γ la quantità di fluido che nell'unità di tempo traversa la curva. Per calcolarlo, dividiamo la curva in archi γ_i mediante i punti $\mathbf{r}_i$. Supponiamo per semplicità che questi archi abbiano tutti la medesima lunghezza l. Supponiamo che la particella che traversa il punto $(x, y) \in \gamma$ mantenga la velocità $\mathbf{V}(x, y)$ anche dopo aver traversato la curva. Se gli archi sono "molto piccoli", la velocità delle singole particelle di liquido che traversano l'arco γ_i si potrà approssimare mediante $\mathbf{V}(x_i, y_i)$. Nell'unità di tempo, queste particelle di liquido riempiono una parte di piano che all'incirca è un parallelogramma la cui base misura l e la cui altezza misura $[\mathbf{V}(x_i, y_i) \cdot \mathbf{n}(x_i, y_i)]$ ove $\mathbf{n}(x_i, y_i)$ è la normale a γ nel punto $\mathbf{r}_i$, si veda la figura 8.5, a destra. La somma

$$\sum_i [\mathbf{V}(x_i, y_i) \cdot \mathbf{n}(x_i, y_i)] l$$

è una somma integrale che per $l \to 0$ approssima

$$\int_\gamma [\mathbf{V}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})] \, \mathrm{d}s \, .$$

Si sceglie quindi quest'integrale come definizione di *flusso* *di un campo vettoriale attraverso una curva.*

Il flusso cambia segno cambiando il verso di percorrenza lungo la curva perché cambiando verso di percorrenza cambia il verso sulla normale.

Supponiamo ora che la curva γ sia semplice e chiusa. In questo caso, usa privilegiare la normale esterna $\mathbf{n_e}$:

$$\mathbf{n_e} = \frac{1}{\sqrt{[x'(t)]^2 + [y'(t)]^2}} \, [y'(t)\mathbf{i} - x'(t)\mathbf{j}]$$

e quindi il flusso uscente dalla regione Ω_γ delimitata da γ è

$$\int_a^b [f(x(t),y(t))y'(t) - g(x(t),y(y))x'(t)]\,\mathrm{d}t = \int_\gamma -g\,\mathrm{d}x + f\,\mathrm{d}y\,.$$

8.3 Analisi vettoriale nel piano

Mostriamo che calcolare l'integrale di certi campi vettoriali equivale al calcolo di integrali multipli. Questo studio va sotto il nome di *analisi vettoriale.*

8.3.1 Una considerazione preliminare

Consideriamo una funzione

$$y = f(x)\,, \qquad x \in [a,b]\,.$$

Questa funzione identifica la curva semplice

$$\gamma_f\,:\; x = t\,, \quad y = f(t)\,, \qquad t \in [a,b]\,.$$

Sia ora $F(x,y)$ una funzione (continua) delle due variabili x ed y e consideriamo l'integrale

$$\int_a^b F(x,f(x))\,\mathrm{d}x\,.$$

Questo è l'integrale

$$\int_{\gamma_f} \mathbf{V}(\mathbf{r})\cdot\,\mathrm{d}\mathbf{r}$$

ove

$$\mathbf{V}(\mathbf{r}) = \mathbf{V}(x,y) = F(x,y)\mathbf{i} + 0\mathbf{j}\,.$$

Col linguaggio delle forme differenziali,

$$\int_a^b F(x,y(x))\,\mathrm{d}x = \int_{\gamma_f} F\,\mathrm{d}x\,.$$

Infatti, $x'(t) = 1$. Analogamente, se

$$x = h(y)\,, \qquad y \in [\alpha,\beta]\,,$$

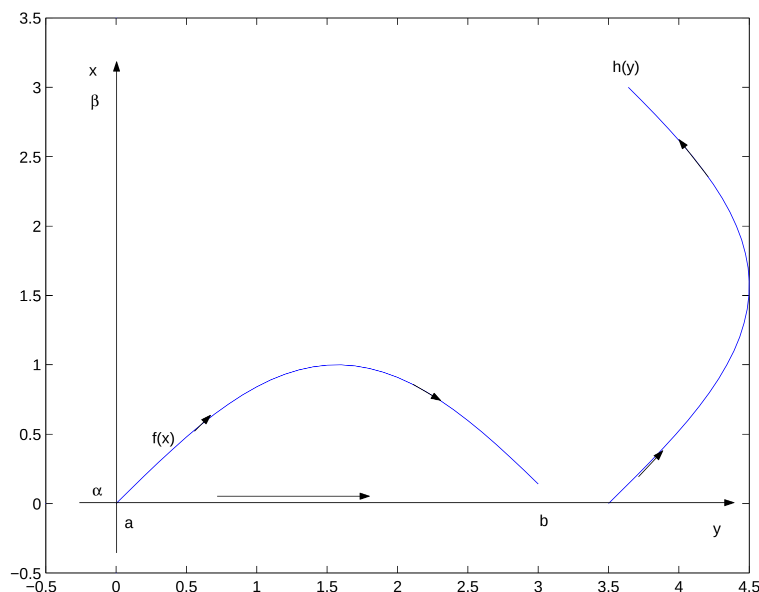
sia γ_h la curva identificata da questo grafico. Si ha

$$\int_\alpha^\beta F(h(y),y)\,\mathrm{d}y = \int_{\gamma_h} \mathbf{W}\cdot\,\mathrm{d}r$$

ove ora $\mathbf{W} = 0\mathbf{i} + F(x,y)\mathbf{j}$.

E' importante esaminare la figura 8.6 e rendersi conto dei versi di percorrenza scelti su queste curve.

Figura 8.6: Versi di percorrenza



8.3.2 Formula di Green

Siano ora γ_f e γ_g due curve cartesiane definite sul medesimo intervallo,

$$y = f(x) \qquad x \in [a,b]$$

e

$$y = g(x) \qquad x \in [a,b]\,.$$

Supponiamo inoltre che per ogni $x \in (a,b)$ sia[5]

$$g(x) > f(x)\,.$$

Indichiamo con Ω la regione delimitata dai due grafici e dai segmenti verticali per $x = a$ e per $x = b$ congiungenti i due grafici. Il segmento si considera nel caso in cui l'uguaglianza non valga in uno dei due estremi. Si veda la figura 8.7, a sinistra.

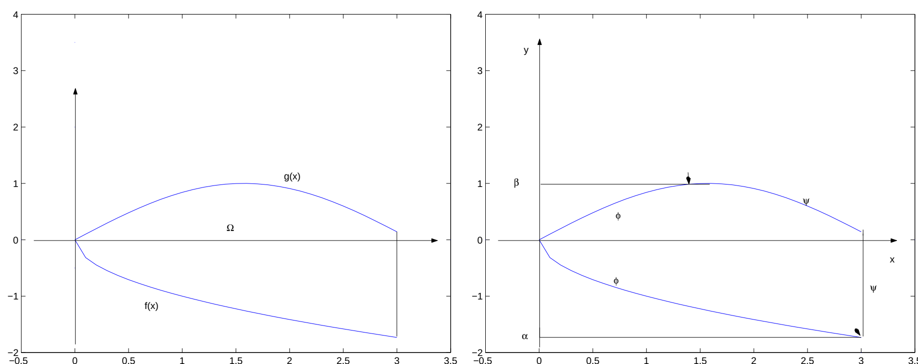
Si noti che la frontiera $\partial\Omega$ di Ω è sostegno di una curva chiusa. Ricordando le nostre convenzioni, indicheremo col medesimo simbolo sia la curva che il suo sostegno; e quindi la indicheremo col simbolo $\partial\Omega$ e, sempre per le nostre convenzioni, $\partial\Omega$ è orientata in senso antiorario.

Vogliamo calcolare

$$\int_\Omega F_y(x,y)\,\mathrm{d}x\ \mathrm{d}y\,.$$

[5]Non si esclude che l'uguaglianza valga negli estremi dell'intervallo.

Figura 8.7: Formula di Green



Quest'integrale si calcola come segue:

$$\int_a^b \left[\int_{f(x)}^{g(x)} F_y(x, y)\, \mathrm{d}y \right] \mathrm{d}x = \int_a^b [F(x, g(x)) - F(x, f(x))]\ \mathrm{d}x = \int_{\gamma_g} \mathbf{V}\, \mathrm{d}\mathbf{r} - \int_{\gamma_f} \mathbf{V}\, \mathrm{d}\mathbf{r}$$

ove

$$\mathbf{V}(x, y) = F(x, y)\mathbf{i} + 0\mathbf{j}\,.$$

Guardando i versi di percorrenza, si vede che

$$\int_\Omega F_y(x, y)\, \mathrm{d}x = \int_{\partial\Omega} [-\mathbf{V}] \cdot \mathrm{d}\mathbf{r}\,. \tag{8.15}$$

Si noti che in questo calcolo il segmento verticale dà contributo nullo.

Sia ora $G(x, y)$ una funzione derivabile. Si vuol calcolare

$$\int_\Omega G_x(x, y)\, \mathrm{d}x\ \mathrm{d}y\,.$$

Proiettando la regione Ω sull'asse delle ordinate si trova un segmento $[\alpha, \beta]$. La frontiera di Ω è ancora grafico di due funzioni, la funzione $\phi(x)$ e la funzione $\psi(x)$, si veda la figura 8.7, a destra. I pallini con la freccia servono ad indicare i punti nei quali i due grafici si congiungono. Il grafico di $\psi(x)$ include il segmento verticale.

Calcoliamo

$$\int_\Omega G_x(x, y)\, \mathrm{d}x\ \mathrm{d}y = \int_\alpha^\beta \left[\int_{\phi(y)}^{\psi(y)} G_x(x, y)\, \mathrm{d}x \right] \mathrm{d}y$$

$$= \int_\alpha^\beta [G(\psi(y), y) - G(\phi(y), y)]\ \mathrm{d}y \tag{8.16}$$

$$= \int_{\gamma_\psi} \mathbf{W} \cdot \mathrm{d}\mathbf{r} - \int_{\gamma_\phi} \mathbf{W} \cdot \mathrm{d}\mathbf{r} = \int_{\partial\Omega} \mathbf{W}\, \mathrm{d}\mathbf{r} \tag{8.17}$$

con

$$\mathbf{W}(x,y) = 0\mathbf{i} + G(x,y)\mathbf{j}\,.$$

Si noti che in questo calcolo il segmento verticale dà contributo non nullo, mentre darebbe contributo nullo un eventuale segmento orizzontale sulla frontiera di Ω.

Combinando i due calcoli precedenti si trovano le due formule seguenti:

$$\int_\Omega [G_x(x,y) - F_y(x,y)]\;\mathrm{d}x\;\mathrm{d}y = \int_{\partial\Omega} [F\,\mathrm{d}x + G\,\mathrm{d}y]\;, \tag{8.18}$$

ottenuta sottraendo la (8.15) con la (8.17).

Sommandole si trova invece

$$\int_\Omega [G_x(x,y) + F_y(x,y)]\;\mathrm{d}x\;\mathrm{d}y = \int_{\partial\Omega} [-F\,\mathrm{d}x + G\,\mathrm{d}y]\;. \tag{8.19}$$

Osservazione importante

Le formule (8.18) e (8.19) valgono anche se la regione con cui si lavora è delimitata da più di due grafici e più in generale essa vale per qualsiasi regione di Jordan Ω_γ, delimitata da una curva γ regolare a tratti. Quindi non c'è più ragione di fare intervenire i trapezoidi delle funzioni $f(x)$ e $g(x)$, e quindi di indicare con lettere maiuscole le funzioni da integrare. Per questo, da ora in poi useremo la notazione più usuale

$$\int_\Omega [g_x(x,y) - f_y(x,y)]\;\mathrm{d}x\;\mathrm{d}y = \int_{\partial\Omega} [f\,\mathrm{d}x + g\,\mathrm{d}y]\;,$$
$$\int_\Omega [g_x(x,y) + f_y(x,y)]\;\mathrm{d}x\;\mathrm{d}y = \int_{\partial\Omega} [-f\,\mathrm{d}x + g\,\mathrm{d}y]\;,$$

con le lettere minuscole.

Le due formule precedenti sono tra loro equivalenti (si passa dall'una all'altra cambiando f in $-f$). Hanno però due diversi significati fisici, e per questo è bene conservarle ambedue. Inoltre, ambedue la formula hanno un equivalente nello spazio $\mathbf{R}^3$, ma in tal caso le due formule che si ottengono non possono ridursi l'una all'altra con una semplice trasformazione.

La formula

$$\int_\Omega [g_x(x,y) - f_y(x,y)]\,\mathrm{d}x\;\mathrm{d}y = \int_{\partial\Omega} f\,\mathrm{d}x + g\,\mathrm{d}y\,. \tag{8.20}$$

si chiama *formula di Green*.

Osservazione 230 Si noti che $[g_x(x,y) - f_y(x,y)]$ è la componente lungo il versore $\mathbf{k}$ del rotore del campo vettoriale $\mathbf{V}(x,y) = f(x,y)\mathbf{i} + g(x,y)\mathbf{j} + 0\mathbf{k}$. La formula (8.20) si scrive quindi come

$$\int_\Omega (\mathrm{rot}\,\mathbf{V})\cdot\mathbf{k}\;\mathrm{d}x\;\mathrm{d}y = \int_{\partial\Omega} \mathbf{V}\cdot\,\mathrm{d}\mathbf{r} = \int_{\partial\Omega} f\,\mathrm{d}x + g\,\mathrm{d}y\,. \tag{8.21}$$

Col linguaggio dei campi vettoriali la formula si interpreta come segue: il flusso del rotore del campo vettoriale $\mathbf{V}(x,y)$ attraverso la superficie piana Ω è uguale alla circuitazione del campo vettoriale stesso lungo $\partial\Omega$. Naturalmente si sottintende che $\partial\Omega$ sia positivamente orientata. ∎

Osservazione 231 Torniamo a considerare la formula

$$\int_{\Omega} [g_x(x,y) + f_y(x,y)] \, \mathrm{d}x \, \mathrm{d}y = \int_{\partial\Omega} [-f \, \mathrm{d}x + g \, \mathrm{d}y] \, . \tag{8.22}$$

Quest'uguaglianza si chiama *Teorema della divergenza nel piano*. Come si è detto, essa è una diversa formulazione della formula di Green. Come tale, esso sotto le medesime condizioni, ma ha una diversa interpretazione fisica, che ora illustriamo.

Consideriamo il campo vettoriale

$$\mathbf{W}(x,y) = g(x,y)\mathbf{i} + f(x,y)\mathbf{j} \, .$$

A destra della (8.22) figura il flusso del campo vettoriale $\mathbf{W}$, uscente dalla regione Ω (regione interna alla curva regolare a tratti $\partial\Omega$). A sinistra figura la divergenza di $\mathbf{W}$. Dunque la (8.22) si scrive anche come

$$\int_{\Omega} \operatorname{div} \mathbf{W}(x,y) \, \mathrm{d}x \, \mathrm{d}y = \int_{\Omega} \nabla \cdot \mathbf{W}(x,y) \, \mathrm{d}x \, \mathrm{d}y = \int_{\partial\Omega} \mathbf{W} \cdot \mathbf{n_e} \, \mathrm{d}s \, . \tag{8.23}$$

Essa si interpretra come segue: il flusso *uscente* da γ del campo vettoriale $\mathbf{W}$ è uguale all'integrale su Ω della divergenza di $\mathbf{W}$.

Quest'interpretazione in particolare spiega l'uso del termine "divergenza". ∎

8.3.3 Formula di Green e forme differenziali

Si è detto che il simbolo " d" si conserva dentro il simbolo di integrale perché aiuta a ricordare le formule. In apparenza ciò non accade per la formula di Green. E' però possibile introdurre dei simboli comodi come segue. Se $f(x,y)$ è una funzione, con $\mathrm{d}f(x,y)$ indichiamo la forma differenziale

$$\mathrm{d}f(x,y) = f_x(x,y) \, \mathrm{d}x + f_y(x,y) \, \mathrm{d}y \, .$$

Se ω è una 1–forma differenziale, definiamo

$$\mathrm{d}\omega = \mathrm{d}\{f \, \mathrm{d}x + g \, \mathrm{d}y\} = f_x \, \mathrm{d}x \, \mathrm{d}x + f_y \, \mathrm{d}y \, \mathrm{d}x + g_x \, \mathrm{d}x \, \mathrm{d}y + g_y \, \mathrm{d}y \, \mathrm{d}y \, .$$

Ora introduciamo le regole seguenti:

$$\mathrm{d}x \, \mathrm{d}x = 0 \, , \qquad \mathrm{d}y \, \mathrm{d}y = 0 \, , \qquad \mathrm{d}y \, \mathrm{d}x = - \, \mathrm{d}x \, \mathrm{d}y \, , \tag{8.24}$$

suggerite dalle regole con le quali si calcola il prodotto vettoriale, (esplicitamente usato nella formula (8.21)):

$$\mathbf{i}\wedge\mathbf{i}=0\,,\qquad \mathbf{j}\wedge\mathbf{j}=0\,,\qquad \mathbf{j}\wedge\mathbf{i}=-\mathbf{i}\wedge\mathbf{j}\,.$$

In questo modo si trova

$$\mathrm{d}\omega=[-f_y+g_x]\,\mathrm{d}x\;\mathrm{d}y \tag{8.25}$$

e questo è proprio l'integrando dell'integrale doppia in (8.20). Dunque, un modo per ricordare la (8.20) è di scriverla come

$$\int_{\partial\Omega}\omega=\int_{\partial\Omega}[f\,\mathrm{d}x+g\,\mathrm{d}y]=\int_{\Omega}\mathrm{d}[f\,\mathrm{d}x+g\,\mathrm{d}y]=\int_{\Omega}\mathrm{d}\omega\,. \tag{8.26}$$

Osservazione 232 Introdurremo più avanti il Teorema di Stokes in $\mathbf{R}^3$, che vale per curve che sono bordo di calotte. Il Teorema di Green è il caso particolare del Teorema di Stokes in $\mathbf{R}^3$ che si ottiene quando la curva è una curva di Jordan e quando la superficie è la regione interna alla curva. ∎

La 2-forma differenziale $\mathrm{d}\omega$ si chiama anche il *differenziale esterno* (o più brevemente il *differenziale*) della 1-forma ω.

8.3.4 Le forme differenziali e le aree piane

Scriviamo la formula di Green scegliendo come forma differenziale la seguente:

$$\omega=-y\,\mathrm{d}x+x\,\mathrm{d}y$$

Si trova

$$\int_{\gamma}-y\,\mathrm{d}x+x\,\mathrm{d}y=\int_{\Omega_\gamma}2\,\mathrm{d}x\;\mathrm{d}y$$

e l'integrale a destra è il doppio dell'area di Ω_γ. Si trova quindi:

Teorema 233 *Sia γ una curva di Jordan* orientata positivamente *e sia Ω_γ la sua regione interna. L'area di Ω_γ è data da*

$$\frac{1}{2}\int_{\gamma}-y\;dx+x\;dy\,.$$

Questa formula è un caso particolare del Teorema di Green e quindi si prova calcolando

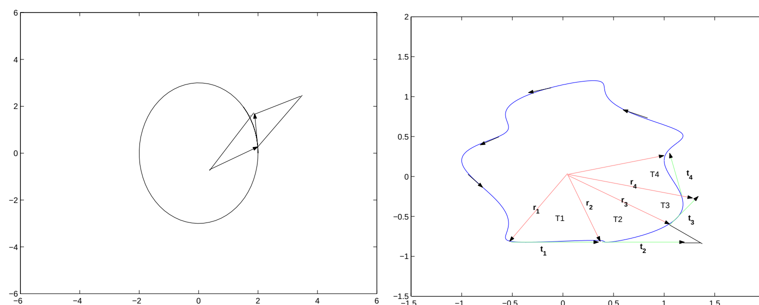
$$\int_{\Omega_\gamma}1\,\mathrm{d}x\;\mathrm{d}y$$

mediante successive integrazioni per parti. E' però interessante vedere un'argomento geometrico che conduce a questa formula.

Sia γ: $t\to\mathbf{r}(t)$, $t\in[a,b]$ un arco piano regolare, semplice e chiuso e sia Ω_γ la sua regione interna.

Per semplicità di esposizione, assumiamo che l'origine degli assi appartenga alla regione interna Ω_γ (caso a cui ci si può sempre ricondurre mediante una traslazione).

Figura 8.8: Area di una regione di Jordan



Si consideri la figura 8.8. Il punto $\mathbf{r}(t)$ varia sull'arco. Dividiamo l'intervallo $[a, b]$ in "piccoli" intervalli, dividendolo con i punti t_i (equidistanti per semplicità). Approssimiamo l'arco $\mathbf{r}(t)$, $t \in [t_i, t_{i+1}]$ col segmento di tangente

$$\mathbf{r} = \mathbf{r}(t_i) + \mathbf{r}'(t_i)(t - t_i)\,, \qquad t \in [t_i, t_{i+1}]$$

come in figura 8.8. Consideriamo i triangoli di vertici l'origine, il punto $\mathbf{r}(t_i)$ e il secondo estremo del segmento di tangente. Si potrebbe provare che quando la finezza della partizione di $[a, b]$ tende a zero, la somma delle aree di questi triangoli tende all'area di Ω_γ, pur di intendere l'area di tali triangoli con segno, in modo da cancellare eventuali parti di area coperte più volte.

L'area (con segno) di ciascuno di questi triangoli è la metà della componente lungo l'asse z del prodotto vettoriale dei vettori $\mathbf{r}(t_i)$ e $\mathbf{r}'(t_i)(t_{i+1} - t_i)$. Per approssimare l'area di Ω_γ dobbiamo quindi sommare la componenti lungo l'asse verticale di

$$\frac{1}{2}\mathbf{r}(t_i) \wedge \mathbf{r}'(t_i)(t_{i+1} - t_i)\,.$$

Si trova così

$$\frac{1}{2}\sum_{i=0}^{N-1}[x(t_i)y'(t_i) - y(t_i)x'(t_i)](t_{i+1} - t_i)\,.$$

Al tendere a zero della finezza della partizione queste somme approssimano

$$\frac{1}{2}\int_a^b [x(t)y'(t) - y(t)x'(t)]\,\mathrm{d}t = \frac{1}{2}\int_\gamma [-y\,\mathrm{d}x + x\,\mathrm{d}y]\,. \tag{8.27}$$

L'argomento precedente non è rigoroso, ma può pienamente giustificarsi.

Chiediamoci ora se l'integrale precedente restituisce l'area di Ω_γ, oppure l'area col segno negativo. Si vede subito dalla figura 8.8 che la somma delle aree dei triangoli viene positiva quando l'orientazione della curva è concorde con quella di $\mathbf{R}^2$; ossia quando la curva è orientata positivamente.

8.3.5 Le estensioni

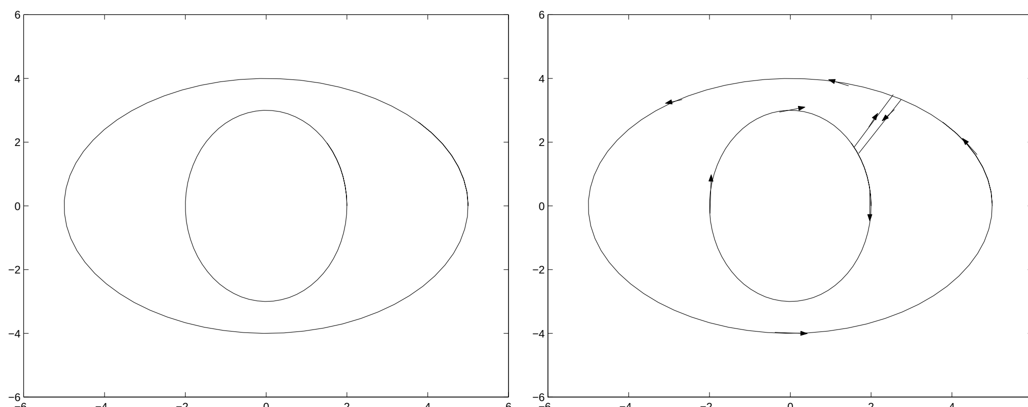
La *formula di Green* è stata provata per una regione molto particolare. Si è già notato che essa vale sotto condizioni molti più generali. E infatti:

Teorema 234 *Le uguaglianze (8.21) e (8.19) valgono per ogni regione Ω_γ, con γ curva regolare a tratti e per ogni coppia di funzioni $f(x,y)$, $g(x,y)$ di classe $C^1(\Omega_\gamma)$, continue sulla chiusura di Ω_γ.*

Mostriamo ora un'estensione importante di queste formule. Quest'estensione è suggerita dalla figura (8.4).

Si tratta di un'estensione di queste formule a regioni delimitate da più curve, come nel caso rappresentato dalla figura 8.9, a sinistra, dove le curve sono due, γ_1 e γ_2, il sostegno dell'una contenuto nella regione interna all'altra.

Figura 8.9: Estensione della formula di Green



Chiamiamo Ω la regione delimitata dalle curve γ_1 e γ_2 e sia $\mathbf{V}(x,y) \in C^1(\Omega)$ un campo vettoriale continuo sulla chiusura di Ω.

Introduciamo due segmenti, S_1 ed S_2, come nella figura a destra, e due piccoli tagli uno su γ_1 e l'altro su γ_2, in corrispondenza a tali segmenti. Si ottiene in questo modo una curva $\tilde{\gamma}$ che verifica sia le condizioni della formula di Green che quelle della formula della divergenza. Vale quindi

$$\int_{\Omega_{\tilde{\gamma}}} [g_x(x,y) - f_y(x,y)]\,\mathrm{d}x\ \mathrm{d}y = \int_{\tilde{\gamma}} f\,\mathrm{d}x + g\,\mathrm{d}y\,,$$
$$\int_{\Omega_{\tilde{\gamma}}} [g_x(x,y) + f_y(x,y)]\,\mathrm{d}x\ \mathrm{d}y = \int_{\tilde{\gamma}} \mathbf{V}\cdot\mathbf{n_e}\,\mathrm{d}s\,.$$

Quando S_1 ed S_2 tendono l'uno all'altro gli integrali lungo i due segmenti si elidono, gli integrali sui due tagli tendono a zero (si ricordino la (8.11) e il Teorema 226.) e l'integrale sulla regione interna a $\tilde{\gamma}$ tende all'integrale su Ω. Si trova quindi

$$\int_\Omega (\operatorname{rot}\mathbf{V})\cdot\mathbf{k}\,\mathrm{d}x\,\mathrm{d}y = \int_{\gamma_2}\mathbf{V}\cdot\,\mathrm{d}\mathbf{r} - \int_{\gamma_1}\mathbf{V}\cdot\,\mathrm{d}\mathbf{r}\,,$$
$$\int_\Omega \operatorname{div}\mathbf{V}\,\mathrm{d}x\,\mathrm{d}y = \int_{\gamma_2}\mathbf{V}\cdot\mathbf{n_e}\,\mathrm{d}s - \int_{\gamma_1}\mathbf{V}\cdot\mathbf{n_e}\,\mathrm{d}s\,.$$

Si noti che il segno negativo viene perchè la curva interna va considerata orientata negativamente, mentre abbiamo convenuto che nel simbolo $\int_\gamma$ la curva chiusa debba essere orientata positivamente.

Nel caso particolare in cui si abbia

$$\operatorname{div}\mathbf{V}(\mathbf{r}) = \nabla\cdot\mathbf{V}(\mathbf{r}) = 0\,,$$

La formula precedente mostra che Se la divergenza di un campo vettoriale è nulla, il flusso del campo vettoriale entrante attraverso γ_1 è uguale a quello uscente attraverso γ_2.

Si discutano risultati analoghi ai precedenti, nel caso in cui la regione delimitata dalla curva più esterna contenga vari "buchi", ciascuno delimitato da una curva di Jordan regolare a tratti.

Osservazione 235 Si osservi che se il campo vettoriale è di classe $C^1(\Omega_{\gamma_1})$ allora le relazioni precedenti si ottengono semplicemente per differenza:

$$\int_\Omega \nabla\cdot\mathbf{V}(\mathbf{r})\cdot\,\mathrm{d}x\,\mathrm{d}y = \int_{\Omega_{\gamma_1}}\nabla\cdot\mathbf{V}(\mathbf{r})\cdot\,\mathrm{d}x\,\mathrm{d}y - \int_{\Omega_{\gamma_2}}\nabla\cdot\mathbf{V}(\mathbf{r})\cdot\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{\gamma_1}\mathbf{W}\cdot\mathbf{n_e}\,\mathrm{d}s - \left[\int_{\gamma_2}\mathbf{W}\cdot\tilde{\mathbf{n}}_{\mathbf{e}}\,\mathrm{d}s\right]\,.$$

In questa formula, $\tilde{\mathbf{n}}_{\mathbf{e}}$ è la normale esterna ad Ω_{γ_2} e quindi interna ad Ω; e γ_2 è percorsa in senso positivo ripetto a Ω_{γ_2} e quindi in senso negativo rispetto a Ω. ∎

8.4 Integrali di superficie

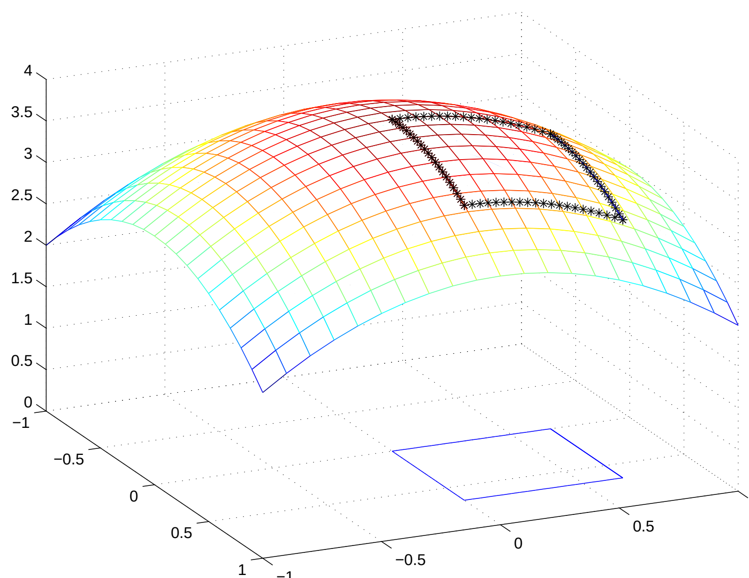
La lunghezza di un arco è stata definita nel paragrafo 6.1.2. Invece, non si è definita l'area di una calotta. Questo è il primo argomento che ora studiamo. Definiremo poi l'integrale su una calotta. Come nel caso delle curve, vedremo che serve definire due tipi diversi di integrali.

8.4.1 Area di una calotta

Consideriamo una calotta $\mathbf{r}(u,v)$ con (u,v) variabili in una regione Ω_γ delimitata da una curva semplice chiusa γ, regolare a tratti. Sia inoltre $\mathbf{r}(u,v)$ di classe $C^1(\Omega_\gamma)$.

Ricordiamo che la chiusura di Ω_γ è un dominio di integrazione. Dunque, è possibile approssimare la regione Ω_γ mediante tanti piccoli rettangoli con i lati paralleli agli assi coordinati, si vedano l'osservazione 204 e il paragrafo 7.1.3. Fissiamo l'attenzione su uno di questi rettangoli e sulla sua immagine sulla superficie (si veda la figura 8.10).

Figura 8.10: Area di una calotta



Per semplicità di notazioni, supponiamo che esso sia un quadrato col vertice a sinistra in basso nel punto $(0,0)$ e che al punto $(0,0)$ corrisponda il punto $(0,0,0)$ (si noti che questo non è il caso illustrato dalla figura).

Consideriamo il lato $(t,0)$ del quadrato, $0 \le t \le \tau$. Ad esso corrisponde una curva $\mathbf{r}(t,0)$ sulla superficie, la cui tangente in $(0,0,0)$ è parallela al vettore $\mathbf{r}_u(0,0)$. Consideriamo il segmento di tangente $\mathbf{r}_u(0,0)t$, $0 \le t \le \tau$.

In modo analogo, a partire dall'altro lato di vertice $(0,0)$, arriviamo a considerare il segmento $\mathbf{r}_v(0,0)t$, $0 \le t \le \tau$. Questi due segmenti (uscenti da $(0,0,0)$) appartengono al piano tangente e in generale non stanno sulla superficie; ma, se τ è "piccolo" il parallelogramma che essi individuano differisce "per poco" dall'immagine del quadrato sulla superficie. Dunque l'area di tale parallelogramma approssima l'area della parte di superficie che viene descritta quando (u,v) varia nel quadrato.

L'area del parallelogramma è $\tau^2 \cdot |\mathbf{r}_u(0,0) \wedge \mathbf{r}_v(0,0)| = \tau^2 |\mathbf{N}(0,0)|$.

Si ripeta questo argomento per ciascuno dei rettangoli e si sommino i risultati. Quello che si trova è una somma di Riemann per l'integrale doppio la funzione $|\mathbf{N}(u,v)|$. Ciò giustifica la definizione seguente:

Definitione 4 Sia

$$(u,v) \to \mathbf{r}(u,v) \qquad (u,v) \in \Omega$$

una superficie regolare. Sia γ una curva semplice e chiusa con sostegno in Ω e sia

$$\Sigma : \qquad (u,v) \to \mathbf{r}(u,v) \qquad (u,v) \in \Omega_\gamma$$

la calotta corrispondente.

Si chiama *area* della calotta Σ il numero

$$\int_{\Omega_\gamma} |\mathbf{N}(u,v)|\,\mathrm{d}u\,\mathrm{d}v\,. \quad \blacksquare$$

Usando la regola per il cambiamento di coordinate degli integrali doppi e il Teorema 202 si potrebbe provare:

Teorema 236 *Due superfici equivalenti, ossia che differiscono solo per la parametrizzazione, hanno la medesima area. Inoltre, l'area non muta cambiando l'orientazione della superficie.*

Se la superficie è data in forma cartesiana,

$$z = f(x,y)\,,$$

la sua area è

$$\int_{\Omega_\gamma} \sqrt{1 + f_x^2(x,y) + f_y^2(x,y)}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Come ulteriore caso particolare, consideriamo il caso di una superficie di rotazione, descritta da

$$z = f(\sqrt{x^2+y^2}) \qquad \omega_\gamma = D_R = \left\{x^2+y^2 \le R^2\right\}\,.$$

In questo caso,

$$\frac{\partial}{\partial x} f(\sqrt{x^2+y^2}) = \frac{x}{\sqrt{x^2+y^2}} f'(\sqrt{x^2+y^2})\,,$$
$$\frac{\partial}{\partial y} f(\sqrt{x^2+y^2}) = \frac{y}{\sqrt{x^2+y^2}} f'(\sqrt{x^2+y^2})$$

e quindi l'area è data da

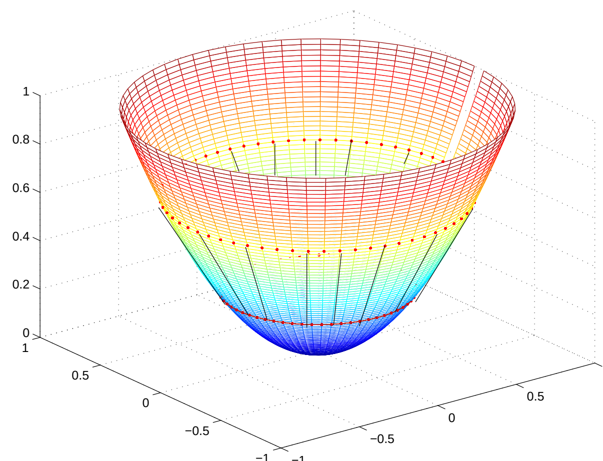
$$\int_{D_R} \sqrt{1 + \left[f'(\sqrt{x^2+y^2})\right]^2}\,\mathrm{d}x\,\mathrm{d}y$$
$$\int_0^{2\pi} \left\{\int_0^R r\sqrt{1+[f'(r)]^2}\,\mathrm{d}r\right\}\mathrm{d}\theta = 2\pi\int_0^R r\sqrt{1+[f'(r)]^2}\,\mathrm{d}r\,.$$

Anche questa formula si chiama *formula di Guldino* e si interpreta facilmente. Dividendo l'intervallo $[0,R]$ mediante i punti r_i, l'integrale si approssima mediante la somma dei numeri

$$(2\pi r_i)\left[(r_{i+1}-r_i)\sqrt{1+[f'(r_i)]^2}\right]$$

e questa è l'area di un tronco di cono "iscritto" nella superficie, come in figura 8.11

Figura 8.11: Calotta di rotazione (caso particolare: paraboloide di rotazione)



8.4.2 Densità superficiale

Si abbia una funzione che ad ogni calotta di una superficie associa un numero. Per esempio, se la superficie è realizzata con una lamiera, la funzione che ad una calotta della superficie associa la sua massa o, nelle applicazioni all'elettrologia, la sua carica. Indichiamo con $m(\Sigma)$ questa funzione. La densità (di massa, di carica ecc.) media su Σ è il numero

$$\rho(\Sigma) = \frac{m(\Sigma)}{A(\Sigma)}\,.$$

In questa formula, $A(\Sigma)$ è l'area della calotta Σ. Vogliamo ora passare dalla densità media alla densità puntuale. Per questo, supponiamo che la superficie sia parametrizzata da

$$(u,v) \to \mathbf{r}(u,v)\,, \qquad (u,v) \in \Omega$$

e supponiamo che la superficie sia semplice, ossia supponiamo che questa trasformazione sia iniettiva.

Fissiamo un punto $\mathbf{r_0}$ sulla superficie,

$$\mathbf{r_0} = \mathbf{r}(u_0,v_0)\,.$$

Sia Q_l un quadrato di lato l il cui centro è (u_0,v_0) e sia $\mathcal{C}_l$ l'immagine del quadrato sulla superficie. Si considera quindi il limite

$$\rho(\mathbf{r_0}) = \lim_{l\to 0} \frac{m(\mathcal{C}_l)}{A(\mathcal{C}_l)}\,.$$

Se questo limite esiste finito, lo chiamiamo la *densità* della funzione m.

In realtà questa definizione è più delicata di quanto possa sembrare perchè per il calcolo del limite si sono privilegiati i quadrati. Niente garantisce che se si fossero scelti dei dischi[6]

[6] o peggio ancora dei rombi che diventano via via più "lunghi e sottili".

si sarebbe trovato il medesimo limite e in effetti generalmente ciò non avviene. Diciamo che, sotto "ipotesi di regolarità" della funzione m ciò non accade. Ma non abbiamo i mezzi necessari per specificare quali siano queste ipotesi.

Il procedimento inverso a questo, porta a definire gli integrali di superficie di prima specia.

8.4.3 Integrali di superfici di prima specie

Così come per le curve, anche sulle superfici si definicono due tipi diversi di integrali. Il primo è suggerito dal problema del calcolo della massa di una superficie, quando si conosce la la densità ρ, che può variare da punto a punto. Ricordando la definizione di densità, per esempio di massa, non meraviglia che per ritrovare la massa totale della calotta che si ottiene quando (u, v) varia in una regione Ω_γ, delimitata da una curva di Jordan γ, si debba calcolare

$$\int_{\Omega_\gamma} \rho(x(u,v), y(u,v), z(u,v))\, |\mathbf{N}(u,v)|\, \mathrm{d}u\ \mathrm{d}v\,.$$

In generale, un integrale della forma

$$\int_{\Omega_\gamma} f(x(u,v), y(u,v), z(u,v))\, |\mathbf{N}(u,v)|\, \mathrm{d}u\ \mathrm{d}v\,. \tag{8.28}$$

(con $f(x, y, z)$ funzione continua definita sulla calotta) si chiama *integrale di superficie di prima* *specie* o semplicemente *integrale superficiale.*

E' chiaro che il valore dell'integrale non cambia cambiando l'orientazione della superficie, dato che nell'espressione dell'integrale figura non direttamente $\mathbf{N}$, ma il suo modulo. Oltre a ciò si potrebbe provare:

Teorema 237 *L'integrale di superficie di prima specie non muta cambiando parametrizzazione.*

Invece di usare la notazione (8.28) si usa in genere la notazione

$$\int_\Sigma f(x,y,z)\,\mathrm{d}\Sigma \qquad \text{oppure} \qquad \int_\Sigma f(x,y,z)\,\mathrm{d}A$$

(A iniziale di "area". Talvolta si scrive $\mathrm{d}S$).

8.4.4 Integrale di superficie di seconda specie

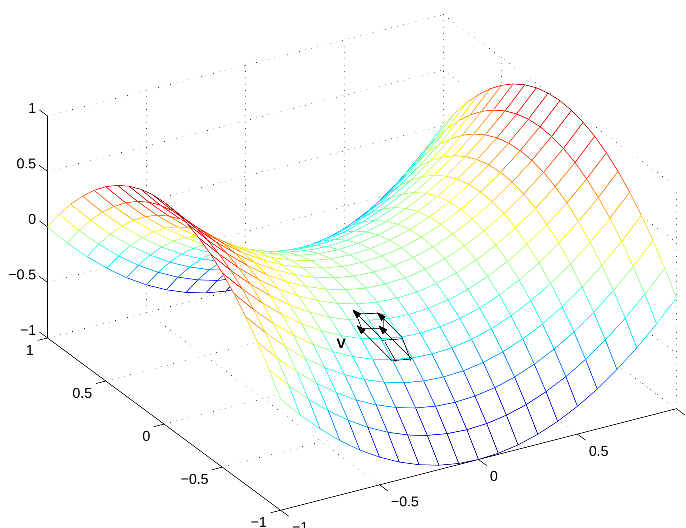
L'*integrale di superficie di* *seconda specie* è suggerito dal calcolo del *flusso* *attraverso una superficie,* che ora definiamo. Sia $\mathbf{V}(x, y, z) = f(x, y, z)\mathbf{i} + g(x, y, z)\mathbf{j} + h(x, y, z)\mathbf{k}$ un campo vettoriale che però ora non interpretiamo come campo di forze. Piuttosto pensiamo che un fluido riempia tutto lo spazio e che la particella che passa per il punto (x, y, z) ci passi con velocità $\mathbf{V}(x, y, z)$. Vogliamo calcolare la quantità di fluido che, nell'unità di tempo, traversa la superficie. Approssimiamo ancora la superficie con tanti piccoli parallelogrammi che giacciono sui piani tangenti, come abbiamo fatto per il calcolo dell'area. Sia $\mathcal{P}$ uno di

essi. La quantità di fluido che nell'unità di tempo lo attarversa è il volume del parallelepipedo che ha $\mathcal{P}$ per base e la cui altezza è, circa,

$$\frac{\mathbf{N}(x_0, y_0, z_0)}{\|\mathbf{N}(x_0, y_0, z_0)\|} \cdot \mathbf{V}(x_0, y_0, z_0)$$

("circa", perché il campo vettoriale non è costante su $\mathcal{P}$). Si veda la figura 8.12.

Figura 8.12: Flusso attraverso una superficie



Il volume va calcolato ora con segno perché non è indifferente che il fluido passi dall'una all'altra parte della superficie, ed è dato da

$$\mathbf{V}(x_0, y_0, z_0) \cdot \mathbf{N}(x_0, y_0, z_0)$$

(si ricordi che $\|\mathbf{N}(x_0, y_0, z_0)\|$ è circa l'area di $\mathcal{P}$ e che $\mathbf{N}(x_0, y_0, z_0)$ è l'area "con segno"). Sommando i contributi di tutti i parallelogrammi si trova una delle somme di Riemann che approssimano l'integrale di $\mathbf{V}(x, y, z) \cdot \mathbf{N}(x, y, z)$. Ciò suggerisce di definire l'*integrale di superficie* *di seconda specie*

$$\int_{\Omega} \mathbf{V}(x(u, v), y(u, v), z(u, v)) \cdot \mathbf{N}(x(u, v), y(u, v), z(u, v))\, \mathrm{d}u\, \mathrm{d}v\,. \tag{8.29}$$

Una notazione più semplice che si usa per indicare questo integrale è

$$\int_{\Sigma} \mathbf{V} \cdot \mathrm{d}\mathbf{\Sigma}\,. \tag{8.30}$$

In questa notazione si sottintende la dipendenza da u e da v.

Si noti che l'integrale di superficie di seconda specie è uguale a

$$\int_\Sigma \mathbf{V}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})\,\mathrm{d}A$$

con $\mathbf{n}(\mathbf{r})$ il versore normale alla superficie.

Si prova:

Teorema 238 *Il valore dell'integrale di superficie di seconda specie non muta cambiando la parametrizzazione della superficie. Esso cambia di segno cambiando l'orientazione.*

Osservazione 239 L'integrale di superficie di seconda specie è un "integrale orientato": esso cambia segno cambiando il verso della normale, ossia l'orientazione della superficie.

Consideriamo ora il caso particolare in cui la superficie Σ è un dominio di integrazione, $\Sigma=\mathcal{D}$, del piano (u,v) e quindi è parametrizzata da

$$x=u\,,\quad y=v\,,\quad z=0\,,\qquad (u,v)\in\Sigma\,.$$

Sia inoltre

$$\mathbf{V}(x,y,z)=\mathbf{V}(x,y)=h(x,y)\mathbf{k}\,.$$

In questo caso, se la superficie è orientata positivamenta

$$\int_\Sigma \mathbf{V}(x,y)\cdot\mathrm{d}\Sigma=\int_\Sigma h(x,y)\,\mathrm{d}x\ \mathrm{d}y=\int_{\mathcal{D}} h(x,y)\,\mathrm{d}x\ \mathrm{d}y,\,,$$

l'usuale integrale doppio di $h(x,y)$. Se però la superficie è orientata negativamente, la normale punta verso il basso e

$$\int_\Sigma \mathbf{V}(x,y)\cdot\mathrm{d}\Sigma=\int_\Sigma h(x,y)\,\mathrm{d}y\ \mathrm{d}x\,,$$

e questo vale

$$-\int_{\mathcal{D}} h(x,y)\,\mathrm{d}x\ \mathrm{d}y\,,$$

l'usuale integrale doppio cambiato di segno. ∎

8.4.5 Integrale di superficie di seconda specie e forme differenziali

Sia nel calcolo del lavoro, integrale di curva di prima specie, che nel calcolo del flusso, integrale di superficie di seconda specie, interviene un campo vettoriale, ma le proprietà fisiche di tali campi sono sostanzialmente diverse. E' quindi utile individuare un formalismo matematico che permetta di distinguere un caso dall'altro e di distinguere anche visivamente se un campo vettoriale si vuol usare per il calcolo di un'integrale di curva o di superficie. Per arrivare a ciò, vediamo una formula esplicita per il calcolo dell'integrale (8.30), più esplicita della formula (8.29). Ricordiamo che

$$\begin{aligned}\mathbf{N}(u,v)&=(x_u\mathbf{i}+y_u\mathbf{j}+z_u\mathbf{k})\wedge(x_v\mathbf{i}+y_v\mathbf{j}+z_v\mathbf{k})\\&=[y_uz_v-z_uy_v]\mathbf{i}+[z_ux_v-x_uz_v]\mathbf{j}+[x_uy_v-y_ux_v]\mathbf{k}\,.\end{aligned}$$

L'espressione esplicita di (8.30) è

$$\int_{\Omega} \{f(x,y,z)[y_u z_v - z_u y_v] + g(x,y,z)[z_u x_v - x_u z_v] + h(x,y,z)[x_u y_v - y_u x_v]\}\, \mathrm{d}u\, \mathrm{d}v \tag{8.31}$$

(la dipendenza di x, y, z da u e da v non si è indicata per brevità).

Se $\mathbf{V}(x,y,z) = f(x,y,z)\mathbf{i} + g(x,y,z)\mathbf{j} + h(x,y,z)\mathbf{k}$ è il campo vettoriale che si vuol integrare su una superficie, indichiamolo col simbolo

$$f\, \mathrm{d}y\, \mathrm{d}z + g\, \mathrm{d}z\, \mathrm{d}x + h\, \mathrm{d}x\, \mathrm{d}y\,. \tag{8.32}$$

Scriviamo inoltre

$$\mathrm{d}x = x_u\, \mathrm{d}u + x_v\, \mathrm{d}v\,, \qquad \mathrm{d}y = y_u\, \mathrm{d}u + y_v\, \mathrm{d}v\,, \qquad \mathrm{d}z = z_u\, \mathrm{d}u + z_v\, \mathrm{d}v\,. \tag{8.33}$$

Usando le "regole di calcolo" (8.24) è ora facile ricostruire l'espressione dell'integrale (8.31). L'integrando si ottiene sostituendo le (8.33) in (8.32) e usando le (8.24) (rispetto ai simboli $\mathrm{d}u$ e $\mathrm{d}v$). Si trova:

$$\begin{aligned}
\mathrm{d}y\ \mathrm{d}x &= y_u z_u\, \mathrm{d}u\ \mathrm{d}u + y_u z_v\, \mathrm{d}u\ \mathrm{d}v + y_v z_u\, \mathrm{d}v\ \mathrm{d}u + y_v z_u\, \mathrm{d}v\ \mathrm{d}v \\
&= (y_u z_v - y_v z_u)\, \mathrm{d}u\ \mathrm{d}v \\
\mathrm{d}z\ \mathrm{d}x &= (z_u x_v - z_v x_u)\, \mathrm{d}u\ \mathrm{d}v \\
\mathrm{d}x\ \mathrm{d}y &= (x_u y_v - x_v y_u)\, \mathrm{d}u\ \mathrm{d}v
\end{aligned}$$

come serve per il calcolo dell'integrale (8.31).

Un'espressione della forma (8.32) si chiama una 2-*forma* *differenziale* (il numero 2 ricorda che si vuole integrarla su una superficie, intuitivamente un "foglietto" di dimensione 2.)

Anche le 2-forme differenziali si indicano con lettere greche minuscole,

$$\omega = f\, \mathrm{d}y\, \mathrm{d}z + g\, \mathrm{d}z\, \mathrm{d}x + h\, \mathrm{d}x\, \mathrm{d}y\,.$$

Completiamo le regole (8.24) imponendo:

$$\begin{array}{lll}
\mathrm{d}x\ \mathrm{d}x = 0\,, & \mathrm{d}y\ \mathrm{d}y = 0\,, & \mathrm{d}z\ \mathrm{d}z = 0\,, \\
\mathrm{d}y\ \mathrm{d}x = -\,\mathrm{d}x\ \mathrm{d}y\,, & \mathrm{d}z\ \mathrm{d}x = -\,\mathrm{d}x\ \mathrm{d}z\,, & \mathrm{d}z\ \mathrm{d}y = -\,\mathrm{d}y\ \mathrm{d}z\,.
\end{array} \tag{8.34}$$

(anche queste regole sono suggerite dalle regole del prodotto vettoriale). Con queste notazioni, la (8.31) si ottiene dalla (8.32) sostituendo x, y, z con $x(u,v)$, $y(u,v)$, $z(u,v)$; usando le regole di calcolo precedenti e quindi integrando su Ω.

8.5 Analisi vettoriale nello spazio

Studiamo ora l'analisi vettoriale nello spazio.

8.5.1 Formula della divergenza e formula di Gauss

L'uguaglianza (8.26) correla le 1-forme differenziali e gli integrali su aree piane. Mostriamo che esiste una formula analoga, che correla gli integrali delle 2–forme differenziali (ossia, i flussi di campi vettoriali) e gli integrali di volume.

Limitiamoci a considerare il caso di regioni Ω convesse. Non è difficile mostrare che la formula che troveremo vale anche se la regione Ω può rappresentarsi come unione o differenza di regioni convesse.

Così come abbiamo fatto per la riduzione degli integrali multipli ad integrali iterati, indichiamo con Ω_z la proiezione sul piano (x, y) di Ω,

$$\Omega_z = \{(x, y) \mid \text{esiste } z \text{ tale che } (x, y, z) \in \Omega\} .$$

Analogamente definiamo Ω_x ed Ω_y. Sia $(x, y) \in \Omega_z$. La retta verticale per (x, y) interseca Ω in un segmento $[\phi(x, y), \psi(x, y)]$. Richiediamo che le funzioni $\phi(x, y)$, $\psi(x, y)$ siano di classe C^1. Proprietà analoghe richiediamo anche per le intersezioni con le rette parallele agli assi delle ascisse e delle ordinate.

Sotto queste condizioni, la regione Ω viene ad essere delimitata da "pezzi" di superfici regolari. In ciascun punto di tali superfici è possibile definire la normale. La normale non sarà generalmente definita nei punti in cui due "pezzi" diversi si congiungono. Per semplicità noi supporremo che in ogni punto di $\partial\Omega$ sia possibile definire la retta normale. Vedremo più avanti come indebolire questa condizione.

Consideriamo un punto $\mathbf{r_0} = (x_0, y_0, \psi(x_0, y_0))$ della superficie superficie $z = \psi(x, y)$. Se un punto $\mathbf{r_1} = (x_1, y_1, z_1)$ di questa retta appartiene ad Ω, tutto il segmento che lo congiunge a $\mathbf{r_0}$ è in Ω perché Ω è convessa. Dunque, è possibile definire la normale entrante e la normale uscente da Ω nel punto $\mathbf{r_0}$.

Applichiamo in ogni punto P di Ω la normale uscente ad Ω in tale punto. Tale vettore normale indichiamo col simbolo $\mathbf{N_e}(P)$ (l'indice "e" indica "normale esterna" o uscente dalla superficie).

Per esempio, fissiamo l'attenzione sulla "parte inferiore" della superficie Ω in Fig. 8.13, parametrizzata da

$$x\mathbf{i} + y\mathbf{j} + \phi(x, y)\mathbf{k} \qquad (x, y) \in \Omega_z .$$

Nel caso specifico si tratta di una sfera e se ne considera la parte "sotto l'equatore". Abbiamo definito la normale

$$\mathbf{N}(x, y) = (\mathbf{i} + \phi_x(x, y)\mathbf{k}) \wedge (\mathbf{j} + \phi_y\mathbf{k}) = -\phi_x(x, y)\mathbf{i} - \phi_y(x, y)\mathbf{j} + \mathbf{k} .$$

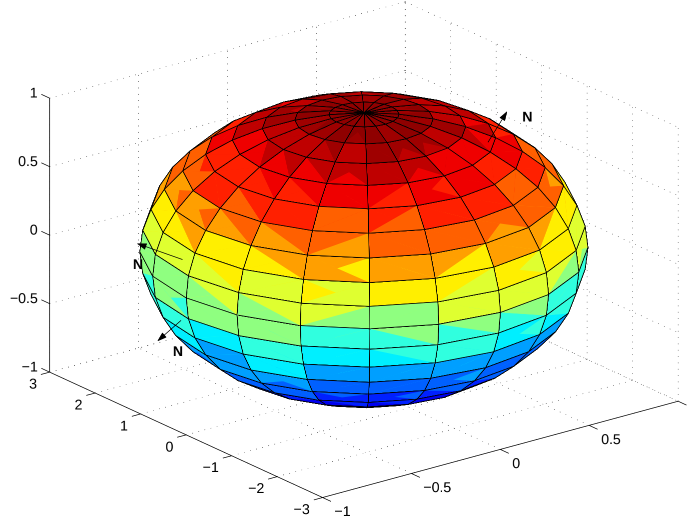
Essendo positivo il coefficiente di $\mathbf{k}$, la normale punta verso l'alto. Dunque entra in Ω. Invece, noi vogliamo la normale che esce da Ω. Dobbiamo quindi cambiare segno alla $\mathbf{N}(x, y)$ e scegliere

$$-\mathbf{N}(x, y) = -(\mathbf{i} + \phi_x(x, y)\mathbf{k}) \wedge (\mathbf{j} + \phi_y\mathbf{k}) = \phi_x(x, y)\mathbf{i} + \phi_y(x, y)\mathbf{j} - \mathbf{k} .$$

Definitione 5 Indicheremo con $\mathbf{N_e}(\mathbf{r})$ la normale nel punto $\mathbf{r}$ alla superficie Σ che racchiude Ω, orientata verso l'esterno di Ω. ∎

Vediamo esplicitamente un esempio.

Figura 8.13: Una sfera e la parte “sotto l'equatore”



Esempio 240 La superficie sia la sfera $x^2 + y^2 + z^2 = 1$. Si tratta quindi di una superficie assegnata in modo implicito. I due emisferi superiore ed inferiore sono parametrizzati rispettivamente da

$$z = \sqrt{1 - x^2 - y^2} \qquad z = -\sqrt{1 - x^2 - y^2}\,, \qquad x^2 + y^2 < 1\,.$$

Si vede facilmente che se si vuole la normale esterna, ossia uscente da Ω, bisogna parametrizzare l'emisfero superiore scegliendo prima x e poi y; nell'emisfero inferiore bisogna scegliere prima y e poi x. ∎

Sia ora $T(x, y, z)$ una funzione derivabile. Calcoliamo

$$\int_\Omega \frac{\partial}{\partial z} T(x, y, z)\, \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = \int_{\Omega_z} \left[\int_{\phi(x)}^{\psi(x)} \frac{\partial}{\partial z} T(x, y, z)\, \mathrm{d}z \right] \mathrm{d}x\ \mathrm{d}y$$
$$= \int_{\Omega_z} T(x, y, \psi(x, y))\, \mathrm{d}x\ \mathrm{d}y - \int_{\Omega_z} T(x, y, \phi(x, y))\, \mathrm{d}x\ \mathrm{d}y\,.$$

La normale esterna alla superficie di Ω, nei punti $(x, y, \psi(x, y))$, rispettivamente $(x, y, \phi(x, y))$, è:

$$\mathbf{N_e}(x_0, y_0, z_0) = -\psi_x(x_0, y_0)\mathbf{i} - \psi_y(x_0, y_0)\mathbf{j} + \mathbf{k}\,,$$
$$\mathbf{N_e}(x_0, y_0, z_0) = \phi_x(x_0, y_0)\mathbf{i} + \phi_y(x_0, y_0)\mathbf{j} - \mathbf{k}$$

Dunque

$$\int_\Omega \frac{\partial}{\partial z} T(x, y, z)\, \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = \int_{\Omega_z} T\,\mathbf{k} \cdot \mathbf{N_e}\, \mathrm{d}x\ \mathrm{d}y\,.$$

Indichiamo con Σ_z le due calotte di parametrizzazione $\phi(x,y)$ e $\psi(x,y)$. Con questa notazione e chiamando $\mathbf{n_e}$ il versore

$$\mathbf{n}_e = \frac{\mathbf{N_e}}{\|\mathbf{n_e}\|}\,,$$

l'ultimo integrale è uguale a

$$\int_{\Sigma_z} T\,\mathbf{k}\cdot\mathbf{n_e}\,\mathrm{d}\Sigma = \int_{\Sigma_z} T\,\mathbf{k}\cdot\,\mathrm{d}\boldsymbol{\Sigma}\,.$$

Si noti che i grafici delle funzioni $\phi(x,y)$ e $\psi(x,y)$ possono non esaurire tutta la frontiera di Ω, ma una parte della frontiera che non appartiene a tali grafici è parte di un cilindro verticale. Ha quindi per normale un vettore ortogonale a $\mathbf{k}$ e quindi non dà contributo al calcolo dell'integrale.

In modo analogo si mostra che

$$\int_\Omega \frac{\partial}{\partial y}S(x,y,z)\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = \int_{\Omega_y} S\,\mathbf{j}\cdot\mathbf{N_e}\,\mathrm{d}x\ \mathrm{d}y = \int_{\Sigma_y} S\,\mathbf{j}\cdot\mathbf{n_e}\,\mathrm{d}\Sigma\,,$$
$$\int_\Omega \frac{\partial}{\partial x}R(x,y,z)\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = \int_{\Omega_x} R\,\mathbf{k}\cdot\mathbf{N_e}\,\mathrm{d}y\ \mathrm{d}z = \int_{\Sigma_x} R\,\mathbf{i}\cdot\mathbf{n_e}\,\mathrm{d}\Sigma$$

(le definizioni di Σ_x e Σ_y sono analoghe a quella di Σ_z).

Sommando i tre integrali si trova

Teorema 241 (della divergenza) *Sia Ω una regione* convessa *e limitata, la cui frontiera è sostegno di una superficie regolare. Vale*

$$\int_\Omega \left[\frac{\partial}{\partial x}R(x,y,z) + \frac{\partial}{\partial y}S(x,y,z) + \frac{\partial}{\partial z}T(x,y,z)\right]\,dx\ dy\ dz$$
$$= \int_\Omega \mathrm{div}\,(R\mathbf{i}+S\mathbf{j}+T\mathbf{k})\ dx\ dy\ dz = \int_\Omega \nabla\cdot(R\mathbf{i}+S\mathbf{j}+T\mathbf{k})\ dx\ dy\ dz$$
$$= \int_{\partial\Omega} \{R\mathbf{i}+S\mathbf{j}+T\mathbf{k}\}\cdot\ d\boldsymbol{\Sigma}\,.$$

La superficie Σ è orientata scegliendo la normale usente da Ω.

Nell'ultimo integrale si intende che la superficie è orientata mediante la normale esterna.

Il teorema della divergenza ha la seguente interpretazione: il flusso di un campo vettoriale uscente da una superficie è uguale all'integrale di volume della divergenza del campo vettoriale stesso.

Il flusso è positivo quando il campo vettoriale "esce" da Ω. In tal caso, l'integrale della divergenza è positivo, e ciò, come si è già notato, spiega l'uso del termine "divergenza".

Consideriamo ora una caso particolare: supponiamo che il campo vettoriale $\mathbf{V}(x,y,z)$ sia il gradiente di una funzione $\phi(x,y,z)$, a valori reali. In questo caso,

$$\mathrm{div}\nabla\phi(x,y,z) = \nabla\cdot\nabla\phi(x,y,z)$$
$$= \phi_{xx}(x,y,z) + \phi_{yy}(x,y,z) + \phi_{zz}(x,y,z) = \Delta\phi(x,y,z)$$

e quindi si trova:

$$\boxed{\int_\Omega \Delta\phi(x,y,z)\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z = \int_{\partial\Omega} \nabla\phi\cdot\,\mathrm{d}\mathbf{\Sigma}\,.}$$

Questo caso particolare del *Teorema della divergenza* va sotto il nome di *Formula di Gauss*.

> In fisica hanno importanza queli campi vettoriali che hanno flusso nullo attraverso ogni superficie chiusa. Tali campi si chiamano *campi solenoidali*. Una condizione sufficiente perchè un campo vettoriale sia solenoidale è che la sua divergenza sia nulla. Questa condizione è solo sufficiente come prova l'esempio seguente:
>
> $$\mathbf{F}(x,y,z) = \frac{x}{(x^2+y^2+z^2)^{3/2}}\mathbf{i} + \frac{y}{(x^2+y^2+z^2)^{3/2}}\mathbf{j} + \frac{z}{(x^2+y^2+z^2)^{3/2}}\mathbf{k}$$
>
> (proporzionale al campo elettrico di una carica puntiforme, il cui flusso attraverso una sfera concentrata nella carica non è nullo, nonostante che $\nabla\cdot\mathbf{F}(x,y,z) = 0$). Se accade che il campo è ovunque definito e di classe C^1 allora il campo è solenoidale se e solo se la sua divergenza è nulla.

Vediamo ora di scrivere il *il Teorema della divergenza* con l'uso del simbolismo delle forme differenziali.

Sia

$$\omega = R\,\mathrm{d}y\;\mathrm{d}z + S\,\mathrm{d}z\;\mathrm{d}x + T\,\mathrm{d}x\;\mathrm{d}y\,.$$

Usando le regole (8.34) si trova

$$\mathrm{d}\omega = (R_x + S_y + T_z)\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z\,. \tag{8.35}$$

Per uniformità, usa chiamare "*3–forma differenziale*" un'espressione della forma

$$f(x,y,z)\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z\,.$$

Essa è niente altro che la funzione scalare $f(x,y,z)$. Si usa questa notazione, per intendere che questa funzione va integrata su una regione dello spazio $\mathbf{R}^3$, con l'avvertenza però che il segno cambia per ogni inversione nell'ordine dei tre simboli dx, dy e dz. Ossia, per esempio,

$$f(x,y,z)\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z = -f(x,y,z)\,\mathrm{d}y\;\mathrm{d}x\;\mathrm{d}z\,,$$
$$f(x,y,z)\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z = f(x,y,z)\,\mathrm{d}y\;\mathrm{d}y\;\mathrm{d}x\,.$$

Con queste notazioni, il Teorema della divergenza si scrive

$$\boxed{\int_{\partial\Omega} \omega = \int_\Omega \mathrm{d}\omega\,. \qquad (8.36)}$$

In questa formula si intende di aver orientato la normale verso l'esterno di Ω.

Estensioni

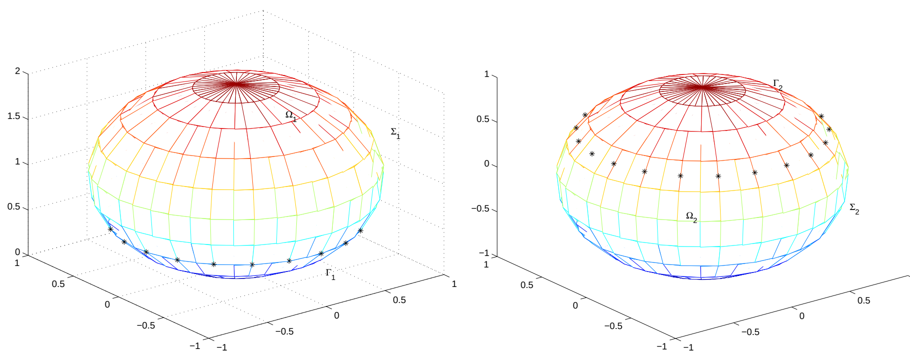
Estendiamo ora la formula della divergenza a regioni più generali. Consideriamo due casi:

Caso 1.

Supponiamo di avere due regioni Ω_1 e Ω_2 su ciascuna delle quali vale il teorema della divergenza, e supponiamo che

- le due regioni non siano contenute l'una nell'altra, ma si intersechino.

Figura 8.14: Le due regioni



Indichiamo con Ω_d e Ω_u rispettivamente l'unione delle due regionie la differenza $\Omega_1 - \Omega_2$:

$$\Omega_d = \{\mathbf{r} \mid \mathbf{r} \in \Omega_1\,,\ \text{e anche } \mathbf{r} \notin \Omega_2\}\,,$$
$$\Omega_u = \{\mathbf{r} \mid \mathbf{r} \in \Omega_1 \quad \text{oppure} \quad \mathbf{r} \in \Omega_2\}\,.$$

Supponiamo che una 2-forma differenziale ω (equivalentemente, un campo vettoriale $\mathbf{V}(\mathbf{r})$) sia di classe $C_1(\mathbf{R}^3)$. Consideriamo la regione Ω_d. In questa regione, la (8.36) diviene:

$$\int_{\Omega_d} \mathrm{d}\omega = \int_{\Omega_1} \mathrm{d}\omega - \int_{\Omega_1 \cap \Omega_2} \mathrm{d}\omega = \int_{\partial\Omega_1} \omega - \int_{\partial(\Omega_1 \cap \Omega_2)} \omega\,. \tag{8.37}$$

Si noti che l'uguaglianza vale perché $\Omega_1 \cap \Omega_2$ è convessa, come unione di regioni convesse. La figura 8.15 mostra una sezione dell'insieme su cui si lavora.

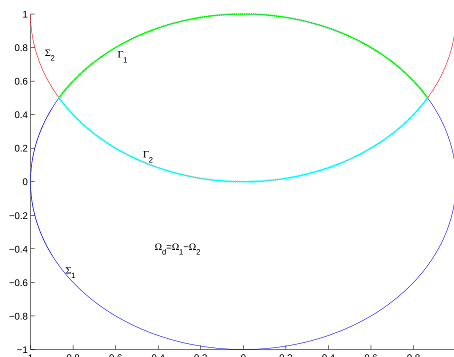
L'insieme $\partial(\Omega_1 \cap \Omega_2)$ è unione di due parti: una parte della frontiera di Ω_1 che chiamiamo Γ_1 e una parte della frontiera di Ω_2 che chiamiamo Γ_2. Chiamiamo invece Σ_1 e Σ_2 le parti rimanenti delle frontiere di Ω_1 e di Ω_2.

Proseguiamo ora le uguaglianze in (8.37). Si ha:

$$\left[\int_{\partial\Omega_1} \omega\right] - \left\{\int_{\partial(\Omega_1 \cap \Omega_2)} \omega\right\} = \left[\int_{\Sigma_1} \omega + \int_{\Gamma_1} \omega\right] - \left\{\int_{\Gamma_1} \omega + \int_{\Gamma_2} \omega\right\}$$
$$\int_{\Sigma_1} \omega + \int_{\Gamma_2} \omega = \int_{\partial(\Omega_1 - \Omega_2)} \omega\,. \tag{8.38}$$

Ciò prova che la formula della divergenza vale anche nella regione non convessa ottenuta come differenza di due regioni convesse.

Figura 8.15: Una sezione della regione



Osservazione 242 Si noti un problema nel calcolo precedente: la frontiera di $\Omega_1 - \Omega_2$ non è regolare e quindi l'uguaglianza in (8.38) non è perfettamente giustificata da quanto abbiamo detto prima. Però la dimostrazione della formula della divergenza si ottiene con una catena di integrazioni per parti, che possono giustificarsi anche nel caso in esame. ▮

Caso 2.

Passiamo ora a considerare la regione $\Omega = \Omega_1 \cup \Omega_2$. Si noti che Ω non è una regione convessa e che

$$\partial\Omega = \Sigma_1 \cup \Sigma_2$$

ove Σ_1 è la parte della frontiera di Ω_1 che non è contenuta in Ω_2; Σ_2 è la parte della frontiera di Ω_2 che non è contenuta in Ω_1.

Sia inoltre

$$\Gamma_1 = [\partial\Omega_1] \cap \Omega_2 \,, \qquad \Gamma_2 = [\partial\Omega_2] \cap \Omega_1 \,.$$

La figura 8.14 mostra, separatamente a sinistra ed a destra, le due regioni con indicate le varie parti della frontiera. Le stelle indicano la curva in cui si intersecano.

La figura 8.16 mostra l'unione delle due regioni, con la parte delle due frontiere "visibile dall'esterno". E' chiaro che

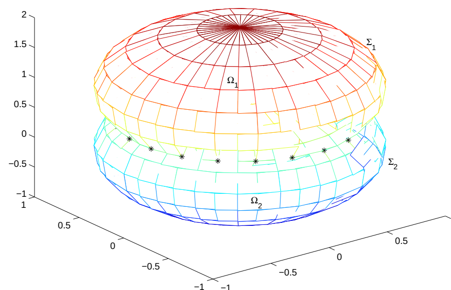
$$\Sigma = \partial\Omega = \Sigma_1 \cup \Sigma_2 \,.$$

Però, né la regione è convessa, né la frontiera è regolare. Ciò nonostante, argomenti analoghi a quelli visti sopra portano a concludere che il teorema della divergenza vale anche nella regione $\Omega = \Omega_1 \cup \Omega_2$.

Osservazione 243 Si potrebbe provare che la classe delle regioni per cui vale il Teorema della divergenza è assai ampia: tutte le regioni limitate la cui frontiera è una superficie regolare *orientabile.* Discuteremo il concetto di orientabilità di una superficie più avanti. Per ora diciamo soltanto che se una superficie regolare ha per sostegno la frontiera di una regione limitata, essa si dice *superficie* *chiusa*[7]. ▮

[7] Non si confonda il termine "superficie chiusa" col termine "insieme chiuso".

Figura 8.16: Ancora la regione



Caso 3.

E' il caso in cui la regione Ω_2 sia contenuta nella regione Ω_1, come può essere il caso di due palle concentriche. In tal caso è facile vedere che la (8.38) vale ancora. Però ora

$$\partial(\Omega_1 - \Omega_2) = [\partial\Omega_1] \bigcup [\partial\Omega_2] \ .$$

La normale va scelta uscente da $\Omega_1 - \Omega_2$. E quindi sulla frontiera di Ω_1 va scelta la normale uscente da Ω_1; sulla frontiera di Ω_2 va scelta la normale entrante in Ω_2.

8.5.2 La formula di Stokes: il caso delle superfici parametriche

La *formula di Stokes* dà un'ulteriore relazione che ora collega integrali di curva nello spazio e integrali di superficie.

Osservazione 244 Abbiamo specificato "curva nello spazio" per intendere che se in queste considerazioni interviene la normale alla curva, come nell'osservazione 245, questa va calcolata con la regola con cui si calcola la normale alla curva nello spazio; ossia, $\mathbf{n}(s)$ è colineare ed ha lo stesso verso di $\mathbf{t}'(s)$. E ciò anche se, per caso, la curva che si considera appartiene ad un piano. ∎

Consideriamo una superficie parametrica Σ

$$\Sigma : \qquad (u,v) \to \mathbf{r}(u,v) = x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k}\,, \qquad (u,v) \in \Omega\,.$$

Supponiamo che la parametrizzazione sia iniettiva e di classe C^2.
Consideriamo una curva semplice e chiusa, con sostegno in Ω:

$$\gamma : \qquad t \to u(t)\mathbf{i} + v(t)\mathbf{j}\,, \qquad t \in [a,b]\,.$$

Indichiamo con Ω_γ la regione interna alla curva γ. La regione Ω_γ è contenuta in Ω e la restrizione di $\mathbf{r}(u,v)$ alla chiusura di Ω_γ definisce una calotta che indichiamo con Σ_1. Diciamo che la calotta Σ_1 ha per *bordo*[8] la curva γ_1 parametrizzata da

$$\begin{aligned} t &\to \mathbf{r}(t) = \mathbf{r}(u(t),v(t)) \\ &= x(u(t),v(t))\mathbf{i} + y(u(t),v(t))\mathbf{j} + z(u(t),v(t))\mathbf{k}\,, \qquad t \in [a,b]\,. \end{aligned}$$

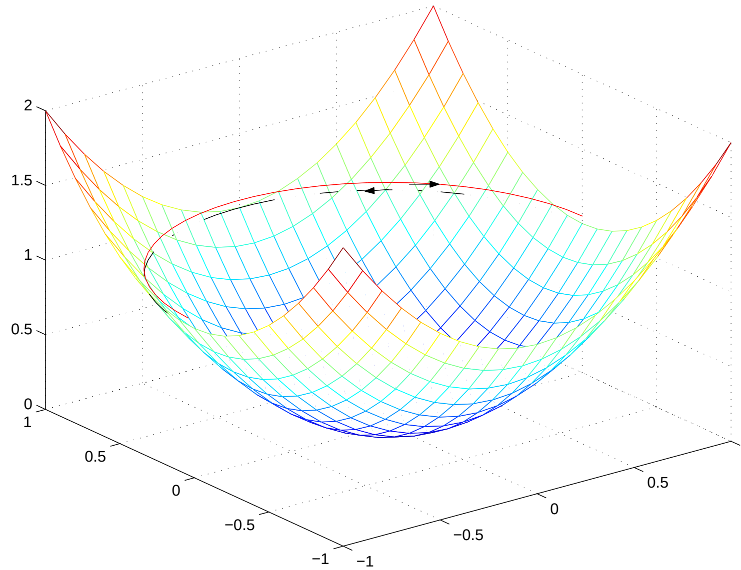
La scelta del verso di percorrenza della curva γ stabilisce un'orientazione su γ_1 mentre la scelta di considerare $\mathbf{i}$ come primo elemento della base di $\mathbf{R}^2$ stabilisce un'orientazione sulla normale a Σ_1. Le due orientazioni non hanno relazioni tra loro: per esempio sia

$$\Sigma : \quad \mathbf{r}(u,v) = u\,\mathbf{i} + v\,\mathbf{j} + \sqrt{1-u^2-v^2}\,\mathbf{k}\,.$$

Le due curve

$$\gamma : \quad t \to \frac{1}{2}[\cos t\,\mathbf{i} + \sin t\,\mathbf{j}]\,, \qquad \tilde{\gamma} : \quad t \to \frac{1}{2}[\sin t\,\mathbf{i} + \cos t\,\mathbf{j}]\,,$$

Figura 8.17: Le due orientazioni del bordo



identificano la medesima calotta, ma le corrispondenti curve γ_1 e $\tilde{\gamma}_1$ hanno orientazioni opposte.

Volendo correlare un'integrale sulla superficie Σ_1 con l'integrale sul suo bordo, dobbiamo correlare le due orientazioni. Per fissare le idee, consideriamo assegnata *l'orientazione della superficie*, ossia il verso positivo della normale e scegliamo di conseguenza quella sulla curva, ma niente vieta di fare il contrario e talvolta questo è effettivamente utile.

[8] Si tratta di una definizione matematicamente non soddisfacente, come si è notato al paragrafo 6.3.1.

Definitione 6 Diciamo che l'orientazione della calotta Σ_1 e quella del suo bordo γ_1 sono *concordi* quando vale la *regola d' Ampère*: una persona in piedi sulla superficie nel verso positivo della normale, vede un punto mobile sulla curva passare dalla sua destra alla sua sinistra. ∎

Osservazione 245 Facendo tendere ad un punto della curva la posizione della persona che osserva il moto, si vede che l'orientazione della calotta e quella del suo bordo sono concordi quando i tre versori della tangente alla curva, della normale alla curva e della normale alla superficie, presi in quest'ordine, formano un sistema di riferimento positivo. Supponiamo che un punto $P(t)$ di γ_1 percorra la curva a partire da $P(a)$, portandosi dietro questo sistema di riferimento. Dato che la parametrizzazione della superficie è iniettiva, il punto $P(t)$ torna alla posizione iniziale quando $t = b$. E alla fine del giro il sistema di riferimento viene a trovarsi nella posizione che aveva all'inizio.

Ricordiamo che in queste considerazioni $\mathbf{n}(s) = \mathbf{t}'(s)/||\mathbf{t}'(s)||$. ∎

Vale:

Teorema 246 (di Stokes) *Valgano le ipotesi dette sopra, in particolare sia iniettiva e di classe C^2 la parametrizzazione della superficie.*

Sia γ_1 il bordo di Σ_1 e Σ_1 e γ_1 abbiano orientazioni compatibili. *Sia $\mathbf{V}(x, y, z)$ un campo vettoriale di classe C^1, definito su Ω. Vale*

$$\int_{\Sigma_1} \operatorname{rot} \mathbf{V} \cdot \, d\mathbf{\Sigma} = \int_{\gamma_1} \mathbf{V} \cdot \, d\mathbf{r} \,. \tag{8.39}$$

Il calcolo che prova questo teorema è in appendice.

Il significato fisico del teorema di Stokes è il seguente: La circuitazione del campo vettoriale lungo il bordo di Σ_1 è uguale al flusso attraverso Σ_1 del rotore del campo vettoriale stesso.

Notiamo che, con le notazioni introdotte per le forme differenziali, anche il *Teorema di Stokes* si scrive

$$\int_{\Sigma_1} \mathrm{d}\omega = \int_{\partial\Sigma_1} \omega \,. \tag{8.40}$$

Infatti, se ω è una 1–forma differenziale i cui coefficienti sono le componenti del campo vettoriale $\mathbf{V}$, è immediato calcolare che $\mathrm{d}\omega$ è quella 2–forma differenziale i cui coefficienti sono le componenti del rotore di $\mathbf{V}$.

Confrontiamo ora le espressioni dei teoremi di Green e di Stokes scritte mediante le formule differenziali. Si vede che essi si possono ambedue scrivere nella forma generale (8.40).

Osservazione 247 Si osservi che nel calcolo l'iniettività della parametrizzazione non si è usata. L'iniettività si è solo usata per l'interpretazione data nell'Osservazione 245. Dunque, la formula (8.39) vale senza quest'ipotesi, pur di intendere che l'integrale a destra sia sostituito da

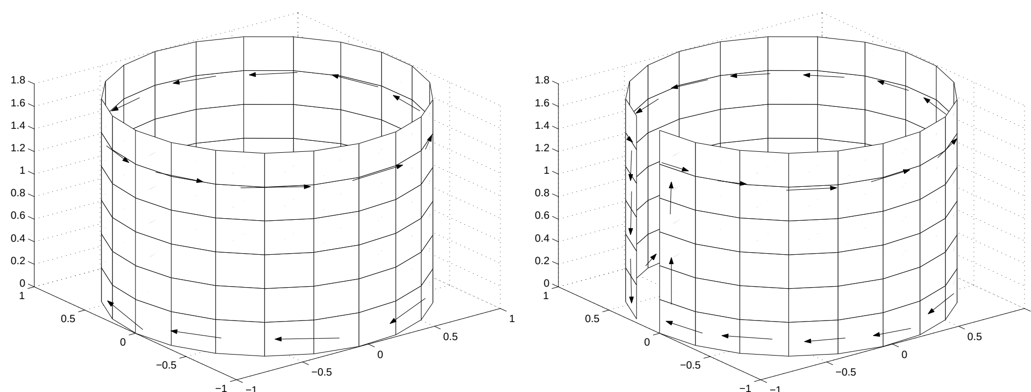
$$\int_a^b \mathbf{V}(x(t), y(t), z(t)) \cdot \mathbf{r}'(t)\,\mathrm{d}t\,.$$

Però in tal caso la curva γ potrebbe essere semplice e la curva γ_1 potrebbe non esserlo; e quindi il valore dell'integrale non dipenderebbe soltanto dagli enti geometrici Σ e γ_1, ma dalla loro parametrizzazione, ciò che non ha senso fisico. Si veda anche l'osservazione 248. ▮

8.5.3 Estensioni

Una prima estensione della formula di Stokes si incontra nel caso in cui la calotta è delimitata da due curve, come nella figura 8.18 a sinistra.

Figura 8.18: Estensione al caso del cilindro



Procedendo come nel caso piano (paragrafo 8.3.5), ossia operando due tagli vicini come nella figura a destra, ci si riconduce al caso che abbiamo già trattato e quindi per una superficie siffatta la formula di Stokes assume l'aspetto

$$\int_\Sigma \mathrm{d}\omega = \int_{\partial\Sigma} \omega = \int_{\gamma_1} \omega - \int_{\gamma_2} \omega\,.$$

Il segno negativo di fronte al secondo integrale dipende dal fatto che la curva γ_2 va orientata in modo discorde rispetto a quello della superficie, si veda la figura.

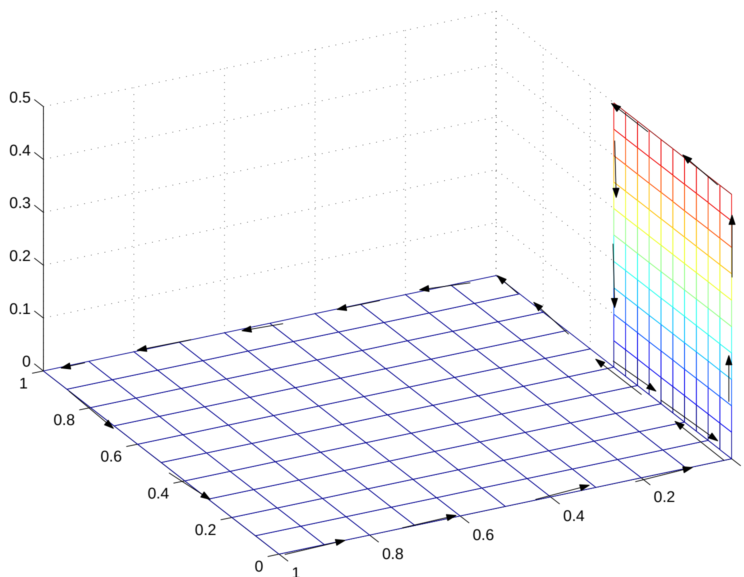
In questo caso diremo che il *bordo* di Σ è costituito dalle due curve γ_1, orientata in modo concorde, e γ_2, orientata in modo discorde, rispetto a Σ.

Si estendono facilmente queste considerazioni al caso in cui il bordo è costituito da più curve.

Ricapitolando, abbiamo introdotto la formula di Stokes per superfici date in forma parametrica. Inoltre, abbiamo esplicitamente supposto che sia possibile orientare la calotta ed

il suo bordo in modo concorde. Il fatto è che ciò talvolta non può farsi e questo è tanto più importante perchè molto spesso in pratica è necessario usare il teorema di Stokes nel caso in cui la superficie è ottenuta "incollando" tanti pezzi di superfici parametriche. Chiameremo *calotta composta* una calotta così ottenuta. Si noti che passando da un pezzo di superficie ad un altro, la normale può variare in modo discontinuo, si veda la figura 8.19

Figura 8.19: Estensione ad una superficie non regolare



E' chiaro dalla figura che

$$\int_{\Sigma} \mathrm{d}\omega = \int_{\Sigma_1} \mathrm{d}\omega + \int_{\Sigma_2} \mathrm{d}\omega$$

e che la formula di Stokes può applicarsi sia a Σ_1 che a Σ_2; che il contributo degli integrali sul lato comune si elide e quindi che l'integrale precedente è uguale alla circuitazione del campo vettoriale sui segmenti che delimitano la superficie, opportunamente orientati. Indicando con γ la curva che ha tali segmenti come sostegno, vale ancora

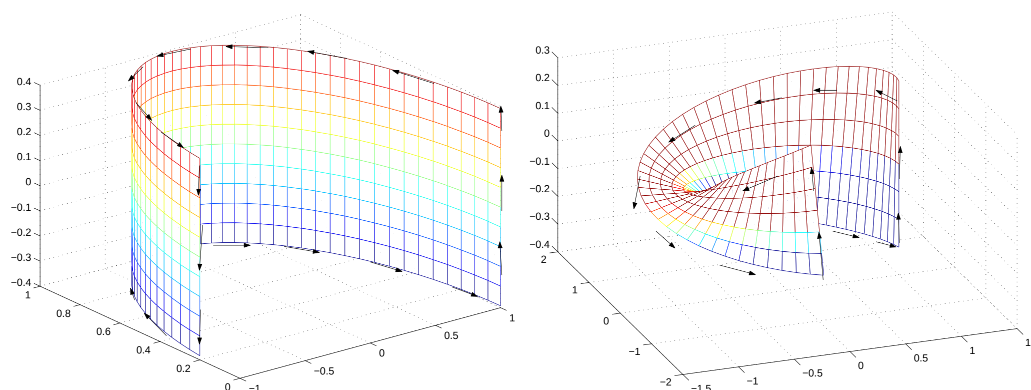
$$\int_{\Sigma} \mathrm{d}\omega = \int_{\partial\Sigma} \omega\,.$$

La discontinuità della normale non costituisce quindi un problema.

Queste considerazioni sono però soltanto apparentemente semplici. Per renderci conto delle difficoltà, applichiamole alle superfici nelle due figure 8.20. Queste sono superfici regolari, il cui bordo è una curva regolare a tratti, e quindi l'applicazione della formula di Stokes non presenta problemi.

Modifichiamo ora le superfici come in figura 8.21. E' ancora ovvio che la formula di Stokes vale. Ma, modifichiamo ulteriormente le superfici, portando i segmenti affiancati a

Figura 8.20: La costruzione del nastro di Möbius



coincidere. Nel caso della superficie a sinistra, la formula di Stokes vale ancora, nella forma generalizzata perché ora il bordo è costituito da due curve. Infatti, gli integrali sui due segmenti si elidono. Invece, tali integrali si sommano nel caso della superficie a destra e quindi per tale superficie, che si chiama *nastro di* *Möbius*, la formula di Stokes non vale.

Osservazione 248 Si noti che nè il cilindro nè il nastro di Möbius sono superfici "semplici", ossia con parametrizzazione iniettiva; ma in un caso la formula di Stokes vale mentre nell'altro non vale; o meglio, vale la formula di Stokes nell'accezione dell'Osservazione 247, che semplicemente significa: contare due volte il contributo dei segmenti che si sovrappongono. Notiamo però che la posizione dei tali segmenti sul nastro di Möbius è arbitraria: possiamo fare un taglio nella posizione che vogliamo e applicare il Teorema di Stokes alla superficie ottenuta. D'altra parte il campo vettoriale cambia da punto a punto e questo vuol dire che il valore dell'integrale cambia a seconda della posizione del taglio, ossia a seconda del modo che usiamo per calcolare l'integrale. Dunque, tale integrale non dipende dalle sole proprietà geometriche del nastro di Möbius, e non ha alcun significato fisico. ∎

Si suggerisce di costruire un modello del nastro di Möbius usando una striscia di carta, e di vedere che il suo bordo consiste di un'unica curva. La circuitazione del campo vettoriale lungo tale curva non ha alcuna relazione col flusso del rotore attraverso il nastro.

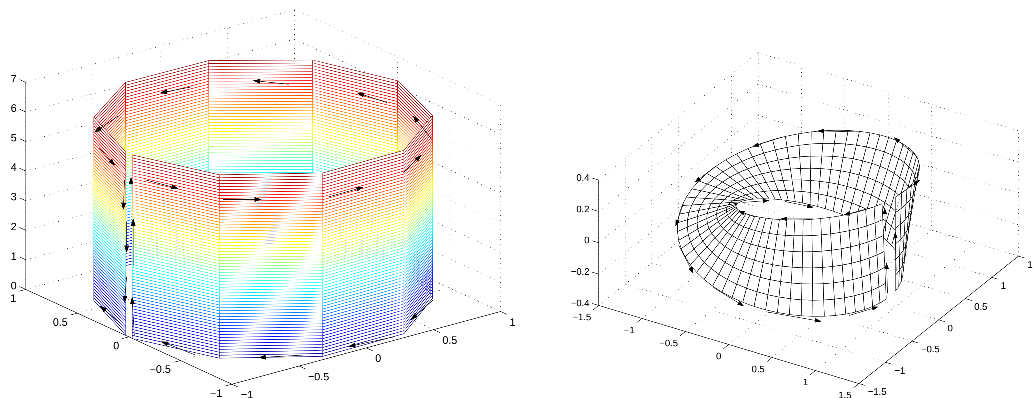
Il nastro di Möbius è una superficie regolare, anche se con parametrizzazione non iniettiva. Dunque, l'unico punto in cui può cedere l'argomentazione che porta alla formula di Stokes è il punto in cui si cerca di orientare la superficie ed il bordo in modo concorde. Infatti, ciò non può farsi. Per vedere ciò, conviene lavorare col versore normale

$$\mathbf{n}(u,v) = \frac{\mathbf{N}(u,v)}{||\mathbf{N}(u,v)||}\,. \tag{8.41}$$

Questo versore è ben definito perchè $\mathbf{N}(\mathbf{r})$ non è mai nullo. Dunque,

$$\det \left[\ \mathbf{r}_u(u,v) \quad \mathbf{r}_v(u,v) \quad \mathbf{n}(u,v)\ \right]$$

Figura 8.21: Il cilindro e il nastro di Möbius



non si annulla mai: o è sempre positivo, o è sempre negativo. Con la scelta fatta per $\mathbf{n}(u, v)$, esso è sempre positivo. Sia ora γ una curva chiusa che gira lungo tutto il nastro di Möbius,

$$\gamma : \quad t \to \mathbf{r}(t)$$

Sia $\mathbf{n}(t) = \mathbf{n}(\mathbf{r}(t))$. La funzione $\mathbf{n}(t)$ varia con continuità, ma la continuità è un concetto locale: niente dice dei valori che $\mathbf{n}(t)$ prende in punti "lontani". Chiediamoci allora cosa accade di $\mathbf{n}(a)$ e di $\mathbf{n}(b)$. Costruendo un modello del nastro di Möbius, si vede immediatamente che

$$\mathbf{n}(a) = -\mathbf{n}(b)\,.$$

E quindi, non è possibile orientare la curva γ in modo concorde alla superficie.

Ciò da una parte spiega come mai la formula di Stokes non vale per il nastro di Möbius e dall'altra conduce alle seguenti definizioni:

Definitione 7 Sia Σ una superficie composta. Essa si chiama una *varietà* se il versore normale $\mathbf{n}(\mathbf{r})$, definito da (8.41), esiste per ogni $\mathbf{r}$, ed è funzione continua di $\mathbf{r}$. La varietà si dice *orientabile* se lungo ogni arco chiuso $\mathbf{r}(t)$, $t \in [a, b]$, si ha

$$\mathbf{n}(\mathbf{r}(a)) = \mathbf{n}(\mathbf{r}(b))\,. \quad \blacksquare$$

La formula di Stokes vale per varietà orientabili, il teorema della divergenza vale in regioni Ω la cui frontiera è una varietà orientabile.

Osservazione 249 Le superfici sono state introdotte parametricamente, ma anche come "superfici di livello" di funzioni $F(x, y, z)$, si veda il capitolo 5. Se la funzione $F(x, y, z)$ è di classe C^1 e il suo gradiente non si annulla, allora le superfici definite da

$$F(x, y, z) = c$$

sono superfici orientabili. ∎

8.6 Appendici

8.6.1 Appendice: fatti da ricordare

Ricordiamo che:

- se
$$\omega = f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z\,, \qquad V = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$$
i coefficienti di $\mathrm{d}\omega$ sono le componenti del rotore di $\mathbf{V}$, ossia di $\nabla \wedge V$.
- Se
$$\omega = f\,\mathrm{d}y\ \mathrm{d}z + g\,\mathrm{d}z\ \mathrm{d}x + h\,\mathrm{d}x\,\mathrm{d}y\,, \qquad V = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$$
allora
$$\mathrm{d}\omega = (\nabla \cdot \mathbf{V})\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z = (\operatorname{div}\mathbf{V})\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\,.$$

Questi fatti vanno ricordati, insieme alla formula seguente, che si ricava facilmente:

$$\operatorname{div}\operatorname{grad} f = \nabla \cdot [\nabla f] = \Delta f = f_{xx} + f_{yy} + f_{zz} = \Delta f\,.$$

8.6.2 Appendice: osservazioni sulla terminologia

Gli argomenti trattati in questo capitolo vanno sotto il nome di "Analisi Vettoriale". I termini usati per indicare i singoli argomenti sono però variabili da autore ad autore (praticamente solo il termine "Teorema di Stokes" è usato da tutti nel medesimo modo, quando applicato alle curve nello spazio). Per esempio, l'integrale di curva di prima specie si chiama anche *integrali curvilineo* mentre l'integrale di curva di seconda specie si chiama anche *integrale di linea*. I due integrali di superficie si chiamano anche, rispettivamente, *integrale superficiale* ed *integrale di flusso*.

Le varie formule integrali che abbiamo incontrato hanno nomi variabili: la formula di Green (nel piano) talvolta si chiama anche *formula di Riemann* ed è niente altro che la particolarizzazione a curve piane della formula di Stokes nello spazio. Il Teorema della divergenza (nello spazio) si chiama anche *Teorema di Gauss* o *formula do Ostrogradski* e talvolta "Formula di Green" ed ha una particolarizzazione al piano che ancora si chiama "formula di Green". E' quindi opportuno memorizzare il significato fisico dei vari teoremi per saperli identificare in contesti diversi.

8.7 Appendice: Una dimostrazione del Teorema di Stokes

Dimostriamo il teorema di Stokes in un caso particolare: supponiamo cioè che la superficie Σ sia cartesiana, parametrizzata da

$$\Sigma : \qquad (x, y) \to (x, y, z(x, y))\,.$$

Però per maggior chiarezza indichiamo con (u, v) i punti di Ω; ossia imponiamo la condizione $x = x(u, v) = u$, $y = y(u, v) = v$.

Notiamo che in questo caso particolare la calotta ed il suo bordo sono orientate in modo concorde quando la curva γ è orientata positivamente; ossia quando una punto mobile su γ vede la regione interna Ω_γ alla sua sinistra.

Sia $\mathbf{V}(x,y,z) = f(x,y,z)\,\mathbf{i} + g(x,y,z)\,\mathbf{j} + h(x,y,z)\,\mathbf{k}$ il campo vettoriale.

Nel nostro caso particolare, si ha

$$\frac{\partial x}{\partial u} = 1\,, \quad \frac{\partial y}{\partial v} = 1\,, \quad \frac{\partial x}{\partial v} = 0\,, \quad \frac{\partial y}{\partial u} = 0\,.$$

Inoltre, $f_x = f_u$, $f_y = f_v$ ed analoga notazione per g ed h. Dunque si calcola:

$$\begin{aligned}
&\int_{\gamma_1} f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z = \int_\gamma \left\{ f\,\mathrm{d}u + g\,\mathrm{d}v + \left(h\frac{\partial z}{\partial u}\,\mathrm{d}u + \frac{\partial z}{\partial v}\,\mathrm{d}v \right) \right\} \\
&= \int_\gamma f\,\mathrm{d}u + g\,\mathrm{d}v + h(z_u\,\mathrm{d}u + z_v\,\mathrm{d}v) = \int_\gamma (f + hz_u)\,\mathrm{d}u + (g + hz_v)\,\mathrm{d}v \\
&= \int_{\Omega_\gamma} \mathrm{d}\,[(f + hz_u)\,\mathrm{d}u + (g + hz_v)\,\mathrm{d}v] \\
&= \int_{\Omega_\gamma} \left\{ \frac{\partial}{\partial u}(g + hz_v) - \frac{\partial}{\partial v}(f + hz_u) \right\} \mathrm{d}u\ \mathrm{d}v \\
&= \int_{\Omega_\gamma} [(g_u + g_z z_u) - (f_v + f_z z_v) + (h_u z_v - h_v z_u)]\ \mathrm{d}u\ \mathrm{d}v\,.
\end{aligned}$$

Mostriamo che anche il flusso del rotore è uguale a questo integrale:

$$\begin{aligned}
&\int_{\Sigma_1} \mathrm{rot}\,\mathbf{V} \cdot \mathrm{d}\mathbf{\Sigma} = \int_{\Omega_\gamma} \{(h_v - g_z)(-z_u) - (f_z - h_u)z_v + (g_u - f_v)\}\ \mathrm{d}u\ \mathrm{d}v \\
&= \int_{\Omega_\gamma} [(g_u + g_z z_u) - (f_v + f_z z_v) + (h_u z_v - h_v z_u)]\ \mathrm{d}u\ \mathrm{d}v\,.
\end{aligned}$$

Ciò prova l'uguaglianza cercata.

Capitolo 9

Campi conservativi

In questo capitolo studieremo l'esistenza del potenziale di un campo vettoriale, equivalentemente, di una primitiva di una forma differenziale. Cambieremo un po' le notazioni perché in molte applicazioni il potenziale si indica col simbolo V. Quindi useremo un altro simbolo, per esempio $\mathbf{F}$, per indicare il campo vettoriale.

9.1 Potenziale

Supponiamo che il campo vettoriale $\mathbf{F}(\mathbf{r})$ definito su una regione Ω sia il gradiente di una funzione[1] $V(\mathbf{r})$ di classe C^1. Si sa dalla fisica che la funzione V si chiama il *potenziale* del campo vettoriale e che un campo vettoriale dotato di potenziale si chiama *conservativo*. I campi conservativi hanno grande importanza per le applicazioni.

> Si faccia attenzione a non confondersi con i segni: in fisica, $V(\mathbf{r})$ si chiama il potenziale del campo $\mathbf{F}(\mathbf{r})$ mentre $-V(\mathbf{r})$ ne è l'*energia potenziale*. In fisica si lavora più frequentemente con l'energia potenziale di $\mathbf{F}(\mathbf{r})$, ossia con la funzione $-V(\mathbf{r})$, talvolta indicata come "energia potenziale $V(\mathbf{r})$". In questo caso $V(\mathbf{r})$ non è il potenziale di $\mathbf{F}(\mathbf{r})$ ma di $-\mathbf{F}(\mathbf{r})$.

[1] ricordiamo che col termine "funzione" si indica sempre una funzione univoca.

Esempio da tener presente per non confondersi

Da un punto di vista fisico, il potenziale è il lavoro fatto **sul corpo** dalla forza esterna; L'energia potenziale è il lavoro che il corpo fa, grazie ad un agente esterno non descritto, **contro la forza esterna,** e quindi **l'energia potenziale è l'energia che il corpo accumula "lavorando contro la forza esterna" e che può restituire in seguito.** Per esempio, se l'asse delle quote è orientato verso l'alto, la forza peso è $-mg$. Il lavoro fatto dalla forza peso in uno spostamento da quota x a quota $x+h$ è $V(h) = -mgh$ e $V_0 + V(h)$ (V_0 costante arbitraria ma fissata) è (un) potenziale della forza. Se lo spostamento è verso l'alto il lavoro è negativo (e quindi il potenziale **diminuisce**) mentre l'energia potenziale del corpo è $E_0 + mgh$ (E_0 costante arbitraria ma fissata) e quindi l'energia potenziale **aumenta** se il corpo viene portato più in alto (da un agente esterno la cui modalità di azione non si descrive: una leva, un razzo o qualunque altro meccanismo che non interessa descrivere).

Vogliamo dare condizioni atte a riconoscere se un assegnato campo vettoriale è conservativo su una regione Ω e, se lo è, vogliamo calcolarne il potenziale.

Per evitare complicazioni puramente tecniche assumeremo che i campi vettoriali e le funzioni siano definiti e con la regolarità che verrà richiesta in una regione $\tilde{\Omega}$ e che la regione Ω in cui si lavora abbia chiusura contenuta in $\tilde{\Omega}$. In questo modo le derivate delle funzioni sono automaticamente continue sia su Ω che sulla sua chiusura.

Col termine "curva" inoltre intenderemo "curva regolare a tratti".

Vediamo subito una condizione necessaria che deve essere soddisfatta dai campi conservativi.

Premettiamo quest'osservazione:

Lemma 250 *Vale*

$$\oint_\gamma \mathbf{F}(\mathbf{r}) \cdot \, d\mathbf{r} = 0$$

per ogni *curva chiusa* γ *di sostegno in* Ω *se e solo se*

$$\int_{\gamma_1} \mathbf{F}(\mathbf{r}) \cdot \, d\mathbf{r} = \int_{\gamma_2} \mathbf{F}(\mathbf{r}) \cdot \, d\mathbf{r}$$

per ogni *coppia di curve* γ_1 *e* γ_2 aventi gli stessi estremi.

Dim. Per la dimostrazione, basta notare che se le due curve γ_1 e γ_2 hanno gli stessi estremi, allora $\gamma_1 - \gamma_2$ è una curva chiusa, si veda il Lemma 228, e quindi

$$0 = \int_{\gamma_1 - \gamma_2} \mathbf{F}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} = \int_{\gamma_1} \mathbf{F}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} - \int_{\gamma_2} \mathbf{F}(\mathbf{r}) \cdot \, \mathrm{d}\mathbf{r} \,. \quad \blacksquare$$

Noto ciò, proviamo:

Teorema 251 *Se* $\mathbf{F}(x, y, z)$ *è un campo conservativo su una regione* Ω *allora*

$$\oint_\gamma \mathbf{F} \cdot \, d\mathbf{r} = 0$$

su ogni curva chiusa di sostegno in Ω. *Equivalentemente,*

$$\int_\gamma \mathbf{F}\cdot\, d\mathbf{r}$$

ha il medesimo valore su tutte le curve γ *di sostegno in* Ω*, che hanno i medesimi estremi.*

Dim. Calcoliamo l'integrale di $\mathbf{F}(\mathbf{r})$ integrale su un arco γ. Sia

$$\mathbf{F}(\mathbf{r}) = V_x(x,y,z)\mathbf{i} + V_y(x,y,z)\mathbf{j} + V_z(x,y,z)\mathbf{k}\,.$$

Si trova

$$\int_\gamma V_x\,\mathrm{d}x + V_y\,\mathrm{d}y + V_z\,\mathrm{d}z =$$
$$\int_a^b [V_x(x(t),y(t),z(t))\dot{x}(t) + V_y(x(t),y(t),z(t))\dot{y}(t) + V_z(x(t),y(t),z(t))\dot{z}(t)]\,\mathrm{d}t$$
$$= \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}V(x(t),y(t),z(t)) = V(x(b),y(b),z(b)) - V(x(a),y(a),z(a))\,.$$

Dunque, in questo caso particolare, l'integrale non dipende dalla curva γ, ma solo dai suoi estremi. In particolare è nullo se gli estremi coincidono, ossia se la curva è chiusa. ∎

Il teorema precedente non è immediatamente usabile, perché richiede infinite verifiche; una per ciascuna curva chiusa di sostegno in Ω. Ciò può lievemente migliorarsi. Notando che ogni curva di sostegno in Ω si può approssimare mediante poligonali, si potrebbe provare:

Lemma 252 *Accade che*

$$\oint_\gamma \mathbf{F}\cdot\, d\mathbf{r} = 0$$

per ogni *curva* γ *di sostegno in* Ω *se e solo se*

$$\oint_P \mathbf{F}\cdot\, d\mathbf{r} = 0$$

per ogni poligonale P *di sostegno in* Ω*.*

Proviamo ora che le condizioni necessarie appena individuate sono anche sufficienti:

Teorema 253 *Un campo vettoriale di classe* C^1

$$\mathbf{F}(\mathbf{r}) = f(\mathbf{r})\mathbf{i} + g(\mathbf{r})\mathbf{j} + h(\mathbf{r})\mathbf{k}$$

ammette potenziale $V(\mathbf{r})$ *se e solo se la sua circuitazione lungo* ogni *poligonale chiusa e semplice di sostegno in* Ω *è nulla.*

Dim. La condizione necessaria si è già provata. Mostriamo che essa è anche condizione sufficiente.

Per fissare le idee, supponiamo $\mathbf{r}\in\Omega\subseteq\mathbf{R}^3$ e quindi

$$\mathbf{F}(\mathbf{r}) = u(x,y,z)\mathbf{i} + v(x,y,z)\mathbf{j} + w(x,y,z)\mathbf{k}\,.$$

Fissiamo un punto $\mathbf{r_0}$ qualsiasi in Ω e costruiamo una funzione $V(\mathbf{r})$ in questo modo: sia $P_{\mathbf{r}}$ una poligonale che congiuge $\mathbf{r_0}$ con $\mathbf{r}$. L'integrale

$$\int_{P_{\mathbf{r}}} \mathbf{F}(\mathbf{c}) \cdot \mathrm{d}\mathbf{c}$$

dipende solo dagli estremi della poligonale, ossia dai punti $\mathbf{r_0}$ e $\mathbf{r}$, dato che l'integrale sulle poligonali chiuse è nullo. Essendo $\mathbf{r_0}$ fissato, il valore dell'integrale dipende solo dal secondo estremo $\mathbf{r}$ della curva. Dunque la funzione

$$V(\mathbf{r}) = \int_{P_{\mathbf{r}}} \mathbf{F}(\mathbf{c}) \cdot \mathrm{d}\mathbf{c}$$

è ben definita. Mostriamo che essa è derivabile e che le sue derivate parziali sono le componenti di $\mathbf{F}$. Consideriamo per questo la derivata rispetto ad x,

$$V_x(x,y,z) = \lim_{hh\to 0} \frac{V(x+h,y,z) - V(x,y,z)}{h} .$$

Calcoliamo $V(x+h,y,z))$ percorrendo prima la curva $P_{\mathbf{r}}$, che congiunge $\mathbf{r_0}$ con $\mathbf{r}$, e poi il segmento parametrizzato da

$$\mathbf{c}_1(t) = x + th\,, \qquad c_2(t) = y\,, \qquad c_3(t) = z\,, \qquad t \in [0,1]\,.$$

Questo segmento congiunge il punto di coordinate (x,y,z) col punto di coordinate $(x+h,y,z)$.

Indichiamo con S questo segmento, così che

$$V(\mathbf{r}+\mathbf{h}) - V(\mathbf{r}) = \left[\int_{\gamma_{\mathbf{r}}} \mathbf{F}(\mathbf{c}) \cdot \mathrm{d}\mathbf{c} + \int_S \mathbf{F}(\mathbf{c}) \cdot \mathrm{d}\mathbf{c}\right] - \int_{\gamma_{\mathbf{r}}} \mathbf{F}(\mathbf{c}) \cdot \mathrm{d}\mathbf{c}$$
$$= \int_S \mathbf{F}(\mathbf{c}) \cdot \mathrm{d}\mathbf{c} = \int_0^1 [u(x+th,y,z)]h \,\mathrm{d}t\,.$$

Dunque, (usando il Teorema 139 nel passaggio dalla penultima all'ultima riga)

$$V_x(x,y,z) = \lim_{h\to 0} \frac{V(x+h,y,z) - V(x,y,z)}{h}$$
$$= \lim_{h\to 0} \frac{1}{h}\left[\int_0^1 [u(x+th,y,z)]h \,\mathrm{d}t\right] = \lim_{h\to 0} \int_0^1 u(x+th,y,z)\,\mathrm{d}t$$
$$= \int_0^1 u(x,y,z)\,\mathrm{d}t = u(x,y,z)$$

come si voleva. In modo analogo si trattano le altre derivate. ∎

Osservazione 254 Ricordiamo che la circuitazione di un campo di forze lungo una curva chiusa si interpreta come il lavoro che il campo compie su un punto che percorre la curva. Il teorema precedente mostra quindi che un campo è conservativo se e solo se esso compie lavoro nullo su ogni punto che percorre una qualsiasi curva chiusa di sostegno in Ω.

Inoltre:

Teorema 255 *Due diversi potenziali del medesimo campo vettoriale, definiti su una medesima regione Ω, hanno differenza costante.*

Dim. Perchè la loro differenza $U(x,y,z) = V_1(x,y,z) - V_2(x,y,z)$ ha derivate parziali tutte nulle. Dunque è costante su ogni poligonale e quindi sulla regione Ω. ∎

L'uso del Teorema 253 per verificare se un campo vettoriale è conservativo, richiede infinite verifiche e quindi non può usarsi per risolvere problemi concreti. Per dare un criterio utilizzabile in pratica, ricordiamo la nostra ipotesi, che il campo vettoriale sia di classe C^1. Quindi il potenziale, se esiste, è di classe C^2 e quindi il *Teorema di Schwarz* relativo all'eguaglianza delle derivate miste mostra:

Teorema 256 *Se il campo vettoriale*

$$\mathbf{F}(\mathbf{r}) = u(x,y,z)\mathbf{i} + v(x,y,z)\mathbf{j} + w(x,y,z)\mathbf{k}$$

di classe C^1 è conservativo, valgono le uguaglianze

$$u_y = v_x\,, \qquad u_z = w_x\,, \qquad v_z = w_y \tag{9.1}$$

in ogni punto di Ω.

Dim. Infatti, sia

$$\mathbf{F}(x,y,z) = V_x(x,y,z)\mathbf{i} + V_y(x,y,z)\mathbf{j} + V_z(x,y,z)\mathbf{k}$$

Ossia, per esempio, $u = V_x$, $v = V_y$. Il Teorema di Schwarz mostra che

$$u_y = V_{x,y} = V_{y,x} = v_x\,.$$

Le altre uguaglianze si ottengono in modo analogo. ∎

Osservazione 257 Le condizioni (9.1) sono state scritte per campi vettoriali su $\mathbf{R}^3$, ma naturalmente valgono anche per campi vettoriali in $\mathbf{R}^2$. Se $n = 2$ queste condizioni si riducono a

$$u_y = v_x\,, \qquad u_x = v_y\,. \quad ∎ \tag{9.2}$$

Le condizioni (9.1) sono le condizioni per avere

$$\operatorname{rot}\mathbf{F}(\mathbf{r}) = \nabla \wedge \mathbf{F}(\mathbf{r}) = 0\,.$$

Un campo vettoriale il cui rotore è nullo si dice *irrotazionale*. Il Teorema 256 si riformula quindi come segue:

Teorema 258 *Ogni campo conservativo è irrotazionale.*

L'esempio seguente mostra che il viceversa non vale:

Esempio 259 Si consideri il campo vettoriale su $\mathbf{R}^2$ dato da

$$\mathbf{F}(x,y) = \frac{-y}{x^2+y^2}\mathbf{i} + \frac{x}{x^2+y^2}\mathbf{j}\,. \tag{9.3}$$

Il campo vettoriale (9.3) è rappresentato nella figura 9.1, a sinistra.

Si vede immediatamente che questo campo vettoriale verifica, ove è definito, le uguaglianze (9.2); però non è conservativo perchè, calcolando la circuitazione lungo la circonferenza parametrizzata da

$$\gamma: \qquad x = r\cos t\,,\quad y = r\sin t\,,\qquad t\in[0,2\pi]$$

si trova

$$r^2\int_\gamma \left(\sin^2 t + \cos^2 t\right)\,\mathrm{d}t = 2r^2\pi \neq 0\,.$$

Dunque, il potenziale non esiste, grazie al Teorema 251.

Se si prova ad usare la costruzione nel Teorema 253, si trova una funzione $V(x,y)$ che però non è estendibile con continuità a tutto il piano privato della sola origine: la funzione che si ottiene non ammette estensione continua ad almeno una semiretta uscente dall'origine. ∎

Osservazione 260 Il campo vettoriale (9.3) è il campo di forze prodotto da un filo percorso da corrente elettrica, in un piano ad esso perpendicolare. Si sa che tale campo di forza può fornire energia ad una particella che è vincolata a percorrere una traiettoria circolare centrata sul filo.

Questo campo di forze è rappresentato nella figura 9.1, a sinistra. ∎

Ricapitolando, abbiamo una condizione necessaria e sufficiente perché un campo sia conservativo, espressa dal Teorema 253. Questo teorema però richiede di fare infinite verifiche, e non è praticamente usabile. Abbiamo poi una semplice condizione necessaria, espressa dal Teorema 256. Però l'esempio precedente mostra che questa condizione non è sufficiente. Essa però diviene sufficiente se la regione su cui si lavora ha una semplice proprietà geometrica:

Definitione 8 Una regione $\Omega \subseteq \mathbf{R}^2$ si dice *semplicemente connessa* se vale

$$\Omega_\gamma \subseteq \Omega$$

per ogni curva di Jordan γ il cui sostegno è in Ω.

Una regione Ω di $\mathbf{R}^3$ si dice *semplicemente connessa* se due qualsiasi punti di Ω possono congiungersi con una curva regolare[2] ed inoltre se ogni curva regolare semplice e chiusa in Ω è bordo di una superficie parametrica semplice il cui sostegno è contenuto in Ω. ∎

[2]questo fatto è automaticamente vero perché Ω è una regione, ossia un aperto connesso. E' stato enunciato esplicitamente per maggior chiarezza.

Intuitivamente, una regione di $\mathbf{R}^2$ è semplicemente connessa quando “non ha buchi”. Una regione di $\mathbf{R}^3$ semplicemente connessa può avere “buchi” che però devono essere “localizzati”. Per esempio, una corona circolare non è semplicemente connessa in $\mathbf{R}^2$ mentre un guscio sferico è semplicemente connesso in $\mathbf{R}^3$. Invece, togliendo da $\mathbf{R}^3$ un cilindro (illimitato in ambedue le direzioni) la regione rimanente non è semplicemente connessa.

Una classe (molto particolare) di insiemi semplicemente connessi in $\mathbf{R}^3$ è quella degli insiemi convessi.

Proviamo ora:

Teorema 261 *Sia* $\mathbf{F}(\mathbf{r})$ *un campo vettoriale di classe* C^1 *su una regione* Ω. *Supponiamo che* $\mathbf{F}(\mathbf{r})$ *sia* irrotazionale.

Se la regione Ω *è* semplicemente connessa *allora il campo è conservativo.*

Dim. Proviamo il teorema in $\mathbf{R}^2$.

Per provare che il campo è conservativo, dobbiamo provare che vale

$$\oint_\gamma \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = 0$$

su ciascuna curva regolare, semplice e chiusa γ di sostegno in Ω. Sia Ω_γ la regione interna a γ.

Dato che la regione Ω è semplicemente connessa, Ω_γ è tutta contenuta in Ω e quindi si può usare il *Teorema di Green,* ossia il *Teorema di Stokes sul piano.*

Si ha quindi

$$\int_\gamma \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \int_{\Omega_\gamma} \operatorname{rot} \mathbf{F}(x,y) \cdot \mathbf{k} \, \mathrm{d}x \, \mathrm{d}y = 0$$

perché il rotore è nullo.

La dimostrazione del teorema in $\mathbf{R}^3$ è analoga: bisogna ricordare che, per ipotesi, ogni curva regolare, semplice e chiusa contenuta in Ω è bordo di una calotta parametrica semplice Σ, tutta contenuta nella regione Ω, sulla quale si può usare il teorema di Stokes in $\mathbf{R}^3$.

Per il Teorema di Stokes, detta γ la poligonale e Σ la calotta[3], vale

$$\int_\gamma \mathbf{F} \cdot \mathrm{d}\mathbf{r} = \int_\Sigma \operatorname{rot} \mathbf{F} \cdot \mathrm{d}\mathbf{\Sigma} = 0 \,.$$

L'esistenza del potenziale segue dall'arbitrarietà della γ, si veda il Teorema 253. ∎

Osservazione 262
- Un disco del piano, o una palla in $\mathbf{R}^3$, sono regioni semplicemente connesse. Dunque ogni campo irrotazionale è localmente conservativo. Difficoltà possono sorgere solamente se ci si “allontana troppo” dal punto di partenza.
- Applicando quest'osservazione al campo vettoriale dell'Esempio 259, possiamo dire che questo campo vettoriale ammette potenziale per esempio in ogni semipiano o in ogni angolo che non contiene l'origine.

[3] orientando Σ e γ con la regola d'Ampère

- La condizione sulla regione Ω è solamente sufficiente. Il potenziale di un campo vettoriale (irrotazionale) può esistere anche in una regione che non è semplicemente connessa, come mostra l'esempio seguente. Il campo vettoriale è definito su $\mathbf{R}^2-(0,0)$:

$$\mathbf{F}(x,y)=\frac{x}{x^2+y^2}\mathbf{i}+\frac{y}{x^2+y^2}\mathbf{j}\,.$$

Questo campo vettoriale ammette potenziale su $\mathbf{R}^2-(0,0)$, dato da

$$V(x,y)=\frac{1}{2}\log(x^2+y^2)\,. \qquad \blacksquare \tag{9.4}$$

9.1.1 Il calcolo del potenziale

Il Teorema 253 insegna a costruire il potenziale di un campo conservativo: basta calcolarne gli integrali lungo curve di forma "semplice", per esempio poligonali che congiungono un punto $\mathbf{r_0}$ fissato col generico punto $\mathbf{r}$ della regione. E' ovvio perè che questa via è praticamente percorribile solamente se due punti della regione possono congiungersi con un segmento, in modo da avere integrali facilmente calcolabili. Il caso più semplice è quello in cui Ω è una regione stellata rispetto ad un punto $\mathbf{r_0}$ (si veda la definizione al paragrafo 3.1.1). In questo caso esiste un punto $\mathbf{r_0}$ che può essere congiunto al generico punto $\mathbf{r}\in\Omega$ mediante un segmento

$$t\to\mathbf{r_0}+t(\mathbf{r}-\mathbf{r_0})\,,\qquad t\in[0,1]\,.$$

In tal caso, Dalla dimostrazione del Teorema 253,

$$V(\mathbf{r})=\int_0^1\mathbf{F}(\mathbf{r_0}+t[\mathbf{r}-\mathbf{r_0}])\cdot[\mathbf{r}-\mathbf{r_0}]\,\mathrm{d}t\,.$$

Peró questa non è l'unica costruzione possibile e non è la più semplice nemmeno nel caso di una regione stellata. Di solito, è più semplice risolvere, con successivi calcoli di primitive, le equazioni

$$V_x(x,y,z)=u(x,y,z)\,,\qquad V_y(x,y,z)=v(x,y,z)\,,\qquad V_z(x,y,z)=w(x,y,z)\,.$$

Vediamo ciò su un esempio.

Esempio 263 Sia

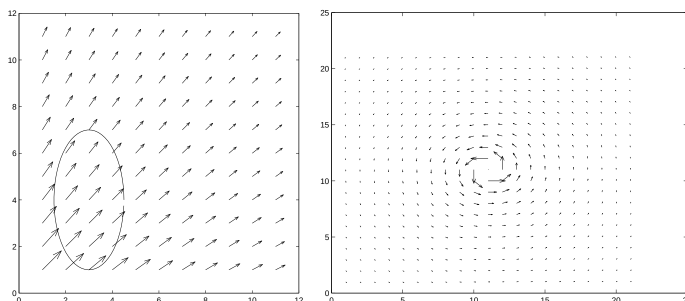
$$\mathbf{F}(x,y,z)=\frac{\mathbf{r}}{\|\mathbf{r}\|^3}\qquad \mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}\,.$$

In questo caso,

$$u(x,y,z)=\frac{x}{(x^2+y^2+z^2)^{3/2}}\,,$$
$$v(x,y,z)=\frac{y}{(x^2+y^2+z^2)^{3/2}}\,,$$
$$w(x,y,z)=\frac{z}{(x^2+y^2+z^2)^{3/2}}\,.$$

Si noti che la funzione non è definita nell'origine; ma sembra di intuire che su ogni curva di Jordan regolare che non passa per l'origine si possa appoggiare una calotta regolare che non incontra l'origine, alla quale applicare il Teorema di Stokes. Inoltre, si vede facilmente

Figura 9.1: Un campo vettoriale conservativo ed uno non conservativo



che il campo vettoriale verifica le uguaglianze (9.1). Si può quindi sperare di costruirne un potenziale in ogni regione semplicemente connessa che non contiene l'origine.

Per questo si noti che integrando rispetto ad x l'uguaglianza

$$V_x(x,y,z) = u(x,y,z) = \frac{x}{(x^2+y^2+z^2)^{3/2}}$$

si trova

$$V(x,y,z) = -\frac{1}{\sqrt{x^2+y^2+z^2}} + \Phi(y,z)\,.$$

Derivando rispetto ad y ed uguagliando a $v(x,y,z)$ si trova

$$\Phi_y(y,z) = 0$$

e quindi $\Phi(y,z)$ non dipende dalla variabile y:

$$\Phi(y,z) = \Phi(z)\,.$$

Derivando ora $V(x,y,z)$ rispetto a z ed uguagliando a $w(x,y,z)$ si trova

$$\Phi'(z) = 0$$

e quindi $\Phi(z)$ viene ad essere costante. il campo vettoriale proposto ammette quindi come potenziali le funzioni

$$V(x,y,z) = -\frac{1}{\sqrt{x^2+y^2+z^2}} + c$$

ove c è una qualsiasi costante.

Si osservi che il campo vettoriale dell'esempio precedente è quello gravitazionale (cambiato di segno) e che il potenziale trovato è il potenziale newtoniano (cambiato di segno). La figura 9.1, a destra, mostra la restrizione del campo vettoriale al piano $x = 0$. ∎

Ricordiamo nuovamente che la condizione di irrotazionalità è necessaria per l'esistenza del potenziale mentre la condizione che la regione sia semplicemente connessa è solamente sufficiente. Un campo irrotazionale potrebbe annettere potenziale anche su una regione non semplicemente connessa. Il potenziale (9.4) mostra un caso di questo tipo.

9.2 Il linguaggio delle 1-forme differenziali

Gli stessi argomenti che abbiamo visto sopra possono riformularsi col linguaggio delle forme differenziali. In tal caso usa una terminologia un po' diversa.

Consideriamo il campo vettoriale e la forma differenziale

$$\mathbf{F}(\mathbf{r}) = f(\mathbf{r})\mathbf{i} + g(\mathbf{r})\mathbf{j} + h(\mathbf{r})\mathbf{k}\,, \qquad \omega = f(\mathbf{r})\,\mathrm{d}x + g(\mathbf{r})\,\mathrm{d}y + h(\mathbf{r})\,\mathrm{d}z\,.$$

Allora:

- la condizione $\nabla \wedge \mathbf{F} = 0$ (campo irrotazionale) equivale alla condizione

$$\mathrm{d}\omega = 0\,. \tag{9.5}$$

  Una forma differenziale che verifica (9.5) si dice *chiusa*.

- la funzione $V(\mathbf{r})$ verifica $\nabla V(\mathbf{r})$ se e solo se

$$\mathrm{d}V(\mathbf{r}) = \omega\,. \tag{9.6}$$

  Una funzione $V(\mathbf{r})$ per cui vale (9.6) si chiama una *primitiva* della 1-forma differenziale ω.

- una forma differenziale dotata di primitive si dice *esatta*.

Possiamo quindi riformulare i teoremi visti per i campi differenziali in questo modo:

- una 1-forma differenziale esatta ha integrale nullo su ogni curva chiusa;
- una 1-forma differenziale esatta è chiusa;
- 1-forma differenziale chiusa su una regione semplicemente connessa è esatta.

Il linguaggio delle forme differenziali è particolarmente comodo quando si vuol passare al caso di forme differenziali di ordine superiore, come ora vediamo.

9.3 Primitive di 2-forme differenziali

Consideriamo ora una 2-forma differenziale

$$\omega = f\,\mathrm{d}y\;\mathrm{d}z + g\,\mathrm{d}z\,\mathrm{d}x + h\,\mathrm{d}x\,\mathrm{d}y\,.$$

In certe applicazioni ha interesse sapere quando esiste una 1-forma differenziale $\tilde{\omega}$ tale che

$$\mathrm{d}\tilde{\omega} = \omega$$

con $\tilde{\omega}$ di classe C^2. Una condizione necessaria è:

$$\mathrm{d}\omega = \mathrm{d}[\,\mathrm{d}\tilde{\omega}] = 0.$$

Infatti, sia

$$\tilde{\omega} = a\,\mathrm{d}x + b\,\mathrm{d}y + c\,\mathrm{d}z\,.$$

Allora, come si è visto,

$$\mathrm{d}\tilde{\omega} = [c_y - b_z]\,\mathrm{d}y\ \mathrm{d}z + [a_z - c_x]\,\mathrm{d}z\ \mathrm{d}x + [b_x - a_y]\,\mathrm{d}x\ \mathrm{d}y\,.$$

Dunque,

$$\begin{aligned}\mathrm{d}[\,\mathrm{d}\tilde{\omega}] &= [c_{yx} - b_{zx}]\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z + [a_{zy} - c_{xy}]\,\mathrm{d}y\ \mathrm{d}z\ \mathrm{d}x + [b_{xz} - a_{yz}]\,\mathrm{d}z\ \mathrm{d}x\ \mathrm{d}y\\ &= [(a_{zy} - a_{yz}) + (b_{xz} - b_{zx}) + (c_{yz} - c_{xy})]\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\end{aligned}$$

e l'ultima espressione è nulla per il Teorema di Schwarz.

La condizione $\mathrm{d}\omega = 0$ si scrive esplicitamente

$$f_x(x,y,z) + g_y(x,y,z) + h_z(x,y,z) = 0\,. \tag{9.7}$$

Una 2-forma differenziale ω che verifica

$$\mathrm{d}\omega = 0$$

si dice *chiusa*.

Se esiste, una 1-forma differenziale $\tilde{\omega}$ per cui

$$\mathrm{d}\tilde{\omega} = \omega$$

si dice una *primitiva* di ω; e una 2-forma differenziale dotata di primitive si dice ancora *esatta*. Proviamo:

Teorema 264 *Una 2-forma differenziale chiusa su un rettangolo è anche esatta.*

Dim. Infatti, supponiamo che la (9.7) valga e mostriamo un modo per costruire la $\tilde{\omega}$. Uguagliando i coefficienti di ω e di $\tilde{\omega}$, si vede che i coefficienti $a(x,y,z)$, $b(x,y,z)$ e $c(x,y,z)$ devono verificare

$$c_y - b_z = f\,, \tag{9.8}$$
$$a_z - c_x = g\,, \tag{9.9}$$
$$b_x - a_y = h\,. \tag{9.10}$$

Proviamo a vedere se si trova una 1-forma $\tilde{\omega}$ che verifica queste uguaglianze e che ha nullo uno dei coefficienti, per esempio il coefficiente c. In tal caso, da (9.8) e (9.8) si trova

$$a(x,y,z) = c_1(x,y) + \int_{z_0}^{z} g(x,y,s)\,\mathrm{d}s$$
$$b(x,y,z) = c_2(x,y) - \int_{z_0}^{z} f(x,y,s)\,\mathrm{d}s$$

dove $c_1(x,y)$ e $c_2(x,y)$ sono arbitrarie funzioni, indipendenti da z.

E ora mostriamo che le funzioni $c_1(x,y)$ e $c_2(x,y)$ si possono determinare in modo che valga anche la (9.10). Per ottenere ciò basta

$$\frac{\partial}{\partial x}c_2(x,y) - \frac{\partial}{\partial y}c_1(x,y) = \Phi(x,y,z) = h(x,y,z) + \int_{z_0}^{z} f_x(x,y,s)\,\mathrm{d}s + \int_{z_0}^{z} g_y(x,y,s)\,\mathrm{d}s\,. \tag{9.11}$$

Notiamo che
$$\frac{\partial}{\partial z}\Phi(x,y,z) = f_x(x,y,z) + g_y(x,y,z) + h_z(x,y,z) = 0$$
e quindi in realtà
$$\Phi(x,y,z) = \Phi(x,y)\,.$$
Notato ciò, si vede che ci sono infiniti modi per soddisfare (9.11). Un modo è di scegliere
$$c_1(x,y) = 0\,, \qquad c_2(x,y) = \int_0^x \Phi(s,y)\,\mathrm{d}s\,. \quad \blacksquare$$

9.4 Alcune formule importanti

Nel corso della trattazione precedente, abbiamo visto che tutte le 0-forme, tutte le 1-forme e tutte le 2-forme (di classe C^2) verificano[4]
$$\mathrm{d}[\,\mathrm{d}\omega] = 0 \tag{9.12}$$

D'altra parte quest'uguaglianza si verifica facilmente anche per le 3-fome (e anzi, per le 3-forme si ha addirittura $\mathrm{d}[f\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z] = 0$. Vogliamo vedere l'aspetto particolare che questa formula assume quando si vuole scrivere per mezzo degli operatori differenziali.

Ricordiamo che:

- se
$$\omega = f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z\,, \qquad V = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$$
i coefficienti di $\mathrm{d}\omega$ sono le componenti del rotore di $\mathbf{V}$, ossia di $\nabla \wedge V$.
- Se
$$\omega = f\,\mathrm{d}y\;\mathrm{d}z + g\,\mathrm{d}z\;\mathrm{d}x + h\,\mathrm{d}x\,\mathrm{d}y\,, \qquad V = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$$
allora
$$\mathrm{d}\omega = (\nabla \cdot V)\,\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z\,.$$

Dunque, la (9.12) assume la forma:
$$\mathrm{rot}\,\mathrm{grad} f = \nabla \wedge [\nabla f] = 0\,,$$
$$\mathrm{div}\,\mathrm{rot}\mathbf{V} = \nabla \cdot [\nabla \wedge \mathbf{V}] = 0\,.$$

Queste formule vanno ricordate insieme a quella, già incontrata e facilmente ricavabile,
$$\mathrm{div}\,\mathrm{grad} f = \nabla \cdot [\nabla f] = \Delta f = f_{xx} + f_{yy} + f_{zz}\,.$$

[4]nel caso delle 0-forme, ossia delle funzioni $f(x,y,z)$, la (9.12) è niente altro che il Teorema di Schwarz.

Capitolo 10

I sistemi di equazioni differenziali

10.1 Introduzione

Ricordiamo dal corso di Analisi matematica 1 che si chiama *equazione differenziale del primo ordine* un'equazione che ha per incognita una funzione $\mathbf{x}(t)$, generalmente a valori vettori, a cui si richiede di verificare

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)) . \tag{10.1}$$

Si parla di problema di Cauchy quando si richiede di risolvere l'equazione differenziale (10.1) insieme all'ulteriore condizione

$$\mathbf{x}(t_0) = \mathbf{x}_0 . \tag{10.2}$$

Convenzionalmente, la variabile t si chiama *tempo*, t_0 si chiama *istante iniziale* e la condizione (10.2) si chiama *condizione iniziale* o *condizione di Cauchy*.

Alcune precisazioni vanno richiamate esplicitamente:

- con le lettere in grassetto si intendono vettori e $\mathbf{x}$ indica un vettore di dimensione n, $n \geq 1$. Quando $n > 1$ la (10.1) si chiama anche un *sistema* di equazioni differenziali.
- Sia $\mathbf{x} \in \mathbf{R}^n$. Il *dominio* dell'equazione differenziale è una regione (quindi un aperto connesso) di $\mathbf{R}^{n+1}$ su cui $\mathbf{f}(t, \mathbf{x})$ è definita. Attenzione che il dominio dell'equazione differenziale è sempre un aperto, anche se $\mathbf{f}(t, \mathbf{x})$ è definita su un chiuso più grande, si veda l'esempio 266.
- la funzione incognita $\mathbf{x}$ è calcolata nel medesimo istante t ovunque essa compare.
- Per *soluzione* dell'equazione (10.1) si intende una funzione $\mathbf{x}(t)$ tale che:
  1. $\mathbf{x}(t)$ è definita su un intervallo (a, b) e ivi derivabile.
  2. per ogni t, la coppia $(t, \mathbf{x}(t))$ appartiene al dominio dell'equazione differenziale.[1]
  3. per ogni $t \in (a, b)$ l'uguaglianza (10.1) è verificata.

[1] si ricordi che per definizione il dominio di un'equazione differenziale è un insieme aperto.

Si dice che la soluzione $\mathbf{x}(t)$ verifica il problema di Cauchy (10.1), (10.2) quando $t_0 \in (a, b)$ e $\mathbf{x}(t_0) = \mathbf{x}_0$.

- Infine, ricordiamo che, come notazione, generalmente invece di (10.1) si scrive

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x})$$

senza indicare la dipendenza di $\mathbf{x}$ da t.

Vedremo che, sotto opportune condizioni sulla funzione $\mathbf{f}(t, \mathbf{x})$, il problema di Cauchy (10.1)-(10.2) ammette soluzione unica, definita su un opportuno intervallo aperto contenente t_0. Scriveremo $\mathbf{x}(t; t_0, \mathbf{x}_0)$ per indicare la soluzione del problema di Cauchy. Quando t_0 è sottinteso, scriveremo $\mathbf{x}(t; \mathbf{x}_0)$.

Si noti che le soluzioni di equazioni differenziali, essendo funzioni continue definite su intervalli, sono curve parametriche.

Per chiarire la definizione di soluzione, consideriamo gli esempi seguenti:

Esempio 265 Sia

$$x' = 1 + x^2 \,.$$

Si sa che $x(t) = \tan t$ verifica l'uguaglianza in ogni punto t in cui la funzione $\tan t$ è definita. Però $x(t) = \tan t$ non è soluzione dell'equazione differenziale perché non è definita su un intervallo. Invece, la restrizione di $\tan t$ all'intervallo $(-\pi/2, \pi/2)$ è soluzione, così come è soluzione la restrizione di $\tan t$ all'intervallo $(\pi/2, 3\pi/2)$.

Consideriamo ora la funzione $\tan(t + \pi/4)$. Questa funzione è soluzione dell'equazione differenziale, sull'intervallo $(-3\pi/4, \pi/4)$. Dunque, soluzioni diverse della medesima equazione differenziale possono avere domini DIVERSI; e in generale una soluzione è definita soltanto su un "piccolo" intervallo. ∎

Inoltre:

Esempio 266 Sia

$$x' = \sqrt{1 - x^2}\,, \qquad x(0) = 0\,.$$

Il dominio della funzione $f(t, x) = f(x) = \sqrt{1 - x^2}$ è l'insieme chiuso $\mathbf{R} \times [-1, 1]$. Questo però non può essere il dominio dell'equazione differenziale perché, per definizione, il dominio dell'equazione differenziale deve essere un aperto. Scegliamo allora come dominio dell'equazione differenziale l'insieme aperto $\mathbf{R} \times (-1, 1)$.

Procedendo per separazione di variabili, si trova che la soluzione è

$$x(t) = \sin t$$

definita su $(-\pi/2, \pi/2)$. Infatti, per $t \to \pm\pi/2$ la soluzione tende rispettivamente a -1 e $+1$. In questi punti, $\sqrt{1 - x^2}$ è definita, ma i punti $(-\pi/2, -1)$ e $(\pi/2, 1)$ non appartengono al dominio dell'equazione differenziale. Dunque, per $t \to \pm\pi/2$, il grafico $(t, x(t))$ della soluzione "esce" dal dominio dell'equazione differenziale. ∎

Chiediamoci ora perché è così importante insistere sul fatto che una soluzione deve rimanere all'interno dell'insieme su cui il secondo membro dell'equazione è definito. La ragione è questa: nei due esempi precedenti la soluzione si trova esplicitamente procedendo per separazione di variabili, ma in pratica la soluzione di un'equazione differenziale deve

calcolarsi numericamente. Consideriamo per esempio la soluzione $x(t) = \sin t$ dell'equazione nell'esempio 266. Se si vuol ridefinire il concetto di soluzione in modo da accettare che la soluzione sia definita anche in $-\pi/2$ e $\pi/2$ allora la soluzione viene ad essere definita su $\mathbf{R}$. Però, nessun metodo numerico riuscirà a trovare tale soluzione perché quando $t = \pi/2$ errori comunque piccoli fanno uscire dal dominio e bloccano l'algoritmo numerico usato per trovare la soluzione, qualunque esso sia.

Come si è visto nel corso di Analisi Matematica 1, si chiama equazione differenziale del secondo ordine un'equazione nell'incognita $x(t)$ del tipo

$$x''(t) = f(t, x(t), x'(t))$$

usualmente scritta senza indicare la dipendenza da t,

$$x'' = f(t, x, x')\,.$$

Si noti che non abbiamo usato lettere in grassetto perché in generale studieremo le equazioni di ordine superiore per incognite scalari (anche se niente vieta di studiare *sistemi* di equazioni di ordine superiore, si veda l'Esempio 269).

Il problema di Cauchy si ottiene ora associando all'equazione differenziale le ulteriori condizioni

$$x(t_0) = x_0\,, \qquad x'(t_0) = x_1$$

ossia ricercando soluzioni dell'equazione differenziale che hanno assegnata posizione e velocità in un medesimo istante.

E' noto dal corso di Analisi Matematica 1 che un'equazione di ordine n è

$$x^{(n)} = f(t, x, x', \ldots, x^{(n-1)})\,. \tag{10.3}$$

Il problema di Cauchy corrispondente si ottiene fissando un punto t_0 e richiedendo che in questo punto la soluzione cercata verifichi le condizioni

$$x(t_0) = x_0\,, \quad x'(t_0) = x_1\,, \ldots, x^{(n-1)}(t_0) = x_{n-1}\,.$$

E' importante notare che ogni equazione differenziale di ordine n può scriversi come sistema di n equazioni differenziali. La (10.3) può scriversi in forma di sistema come segue: si scriva $x_1(t)$ invece di $x(t)$ e quindi si definiscono $x_2(t), \ldots, x_n(t)$ dalle uguaglianze

$$\begin{aligned} &\dot{x}_1(t) = x_2(t) \\ &\dot{x}_2(t) = x_3(t) \\ &\vdots \\ &\dot{x}_n(t) = f(t, x_1(t), x_2(t), \ldots, x_{n-1}(t))\,. \end{aligned} \tag{10.4}$$

Osservazione 267 Si noti l'uso intercambiabile dell'apostrofo e del punto per indicare la derivata rispetto al tempo t.

Mostriamo ora due esempi.

Abbiamo detto che ogni equazione differenziale di ordine n può scriversi come sistema. Invece, il viceversa non vale: esistono sistemi di n equazioni differenziali che **non** corrispondono ad equazioni differenziali di ordine n.

Esempio 268 Si consideri il sistema di equazioni differenziali

$$x' = x\,, \qquad y' = y\,. \tag{10.5}$$

Le sue soluzioni sono $x(t) = e^t x_0$, $y(t) = e^t y_0$. In particolare, la (10.5) è risolta da

$$x(t) = e^t\,, \qquad y(t) = 2e^t\,. \tag{10.6}$$

Si vede subito che nessuna equazione scalare del secondo ordine

$$x'' = f(x, x') \tag{10.7}$$

scritta in forma di sistema del primo ordine, ossia scritta come

$$x' = y\,, \qquad y' = f(x, y)$$

può avere le medesime soluzioni di (10.5). Infatti la prima uguaglianza $x' = y$ non vale se $x(t)$ ed $y(t)$ sono le funzioni in (10.6). ∎

Esempio 269 Si considerino due punti materiali di massa M_1 ed M_2 vincolati a scorrere (senza attrito) su un segmento orizzontale. Il punto M_1 è attaccato ad una molla (di costante elastica k_1) a sua volta fissata nell'origine ed il punto M_2 è collegato al punto M_1 attraverso una seconda molla di costante elastica k_2.

Indichiamo con x ed y rispettivamente l'ascissa del primo e del secondo punto. Allora, le equazioni di Newton per questo sistema sono

$$\begin{aligned} M_1 x'' &= -k_1 x + k_2(y - x) \\ M_2 y'' &= -k_2(y - x)\,. \end{aligned}$$

Si ha quindi un sistema di due equazioni del secondo ordine. Questo può rappresentarsi come sistema di quattro equazioni del primo ordine,

$$\begin{aligned} x_1' &= \tfrac{1}{M_1} x_2 \\ x_2' &= -k_1 x_1 + k_2(y_1 - x_1) \\ y_1' &= \tfrac{1}{M_2} y_2 \\ y_2' &= -k_2(y_1 - x_1)\,. \end{aligned}$$

La fisica insegna che la posizione e velocità “iniziali”, ossia assegnate ad un certo istante di tempo, e le forze cui il sistema è soggetto, determinano l'evoluzione futura del sistema. Dunque, se vogliamo un primo controllo che il modello matematico che abbiamo costruito effettivamente corrisponda alle proprietà fisiche del sistema, dovremo verificare che l'evoluzione nel tempo delle funzioni $x(t)$, $x'(t)$, $y(t)$, $y'(t)$ sia univocamente determinata dalla conoscenza delle costanti M_1, M_2, k_1, k_2 e dei quattro numeri $x(t_0)$, $x'(t_0)$, $y(t_0)$, $y'(t_0)$. A questo quesito risponde il Teorema di Cauchy, Teorema 270. ∎

I sistemi di equazioni differenziali considerati negli esempi precedenti sono tutti *sistemi lineari,* ossia hanno forma

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}$$

con A matrice (generalmente dipendente dal tempo, $A = A(t)$) ed $\mathbf{f} = \mathbf{f}(t)$ dipendente dal solo tempo t e non da $\mathbf{x}$ (anzi, negli esempi, $\mathbf{f} = 0$). Quando $\mathbf{f} = 0$ il sistema si dice *lineare*

omogeneo, altrimenti si chiama *affine.* La funzione $\mathbf{f}(t)$ si chiama *termine noto.* Sono invece esempi di sistemi non lineari i seguenti

$$x' = -x(a+by)\,, \qquad y' = -y(cx+d)$$

che si incontra in problemi di dinamica di popolazioni (equazione di *Lotka-Volterra*) oppure

$$x'' + ax' + cx + \epsilon x^3 = \cos\omega t\,, \quad \text{ossia} \quad \begin{cases} x' = y \\ y' = -ay - cx - \epsilon x^3 + \cos\omega t \end{cases}$$

(equazione *di Duffin*) che si incontra nello studio delle oscillazioni elastiche.

Un altro esempio importante di equazione differenziale non lineare è l'*equazione di van der Pol*

$$x'' = -x + \mu(x^2-1)x'$$

che si incontra nello studio di certi circuiti elettrici.

Consideriamo ora un caso particolare: supponiamo che l'equazione differenziale (10.1) abbia forma

$$\mathbf{x}' = \mathbf{f}(\mathbf{x})$$

con $\mathbf{f}$ che non dipende da t. In questo caso si dice che il sistema è *autonomo* o *tempo invariante.* Esso descrive fenomeni la cui legge fisica è costante nel tempo e le sue soluzioni hanno una proprietà importante: sia $\mathbf{x}(t)$ una soluzione e sia $\mathbf{y}(t) = \mathbf{x}(t+\tau)$ con τ numero fissato ma qualsiasi. Allora,

$$\mathbf{y}'(t) = \mathbf{x}'(t+\tau) = \mathbf{f}(\mathbf{x}(t+\tau)) = \mathbf{f}(\mathbf{y}(t))\,,$$

ossia $\mathbf{y}(t)$ è ancora soluzione. Si osservi che ciò non avviene se $\mathbf{f}$ dipende anche da t perché in questo caso si trova

$$\mathbf{y}'(t) = \mathbf{f}(t+\tau, \mathbf{y}(t))$$

che è una diversa equazione differenziale.

La funzione $t \to \mathbf{y}(t) = \mathbf{x}(t+\tau)$ è una curva ottenuta riparametrizzando (in modo molto semplice) la curva $t \to \mathbf{x}(t)$. Per questa ragione, quando si studiano sistemi autonomi, conviene studiare sia il grafico $(t, \mathbf{x}(t))$ della soluzione sia la "curva soluzione" $t \to \mathbf{x}(t)$. Nel contesto dei sistemi autonomi questa si chiama *orbita*. Graficamente se ne rappresenta il sostegno, ossia l'immagine della funzione e, quando serve, si indica anche il verso di percorrenza al crescere del tempo t.

Un punto $\mathbf{x_0}$ per cui $\mathbf{f}(\mathbf{x_0}) = 0$ è detto *punto di* *equilibrio* o *stazionario* (talvolta si dice anche *soluzione stazionaria*.) In tal caso, $\mathbf{x}(t) \equiv \mathbf{x_0}$ è soluzione costante di (10.1). Se l'equazione differenziale è autonoma, l'orbita di tale soluzione si riduce al punto $\mathbf{x_0}$ stesso. Un punto critico si dice *isolato* se in un suo intorno non vi sono altri punti critici.

Concludiamo dicendo che le equazioni differenziali che stiamo studiando si dicono "ordinarie" in opposizione alle "equazioni differenziali a derivate parziali" che si incontrano quando l'incognita dipende da più variabili e quindi le derivate che compaiono nell'equazione sono derivate parziali. Come esempio di equazione a derivate parziali si consideri

$$\frac{\partial}{\partial t}x + \frac{\partial}{\partial s}x = 0$$

e si provi che tutte le funzioni

$$x(t,s) = \phi(t-s)$$

con ϕ di classe C^1 risolvono quest'equazione.

10.2 Esistenza e unicità di soluzione

Consideriamo il problema di Cauchy

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \qquad \mathbf{x}(t_0) = \mathbf{x}_0 . \tag{10.8}$$

Come si è notato all'esempio 269, vanno conosciute condizioni che assicurino la risolubilità di questo problema, e condizioni che implichino l'unicità della soluzione.

Ricordiamo che per definizione la soluzione deve essere definita in un intervallo che contiene t_0 al suo interno. E' sufficiente però mostrare che la soluzione esiste in $[t_0, t_0 + a)$ perché lo stesso argomento si potrà poi applicare all'intervallo $(t_0 - a, t_0]$. Inoltre, proveremo l'esistenza di soluzioni in un opportuno intervallo, non necessariamente il più grande possibile.

Se $\mathbf{x}(t)$ risolve (10.8) allora $\mathbf{x}(t)$ è una curva parametrica che nel punto $\mathbf{x}(t_0)$ ha per tangente la retta di parametrizzazione

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \mathbf{f}(t_0, \mathbf{x}(t_0))(t - t_0) . \tag{10.9}$$

Ciò suggerisce di costruire come approssimazione della soluzione una spezzata costituita da segmenti delle rette di forma (10.9). Bisogna però usare un po' di cautela, perché si sa già che in generale una soluzione di (10.8) può solo definirsi in un "piccolo" intervallo, si ricordi l'Esempio 265.

Il teorema che si può provare è il seguente:

Teorema 270 (di Cauchy) *Sia $\mathbf{x} = (x_1 , \ldots , x_n)$ e supponiamo che la funzione $\mathbf{f}(t, \mathbf{x})$ e le sue derivate parziali rispetto alle componenti di $\mathbf{x}$ siano continue in una regione D contenente il "rettangolo" R*

$$R = \{(t, \mathbf{x}) \mid t_0 - a \le t \le t_0 + a\, ,\ ||\mathbf{x} - \mathbf{x_0}|| \le b\} .$$

Il problema di Cauchy (10.8) ammette soluzione unica su un intervallo $(t_0 - T, t_0 + T)$ ove T dipende sia da t_0 che da x_0.[2]

La dimostrazione del teorema è piuttosto tecnica, ma è opportuno accennare all'idea che guida la dimostrazione, perché questa suggerisce un modo importante di pensare alle equazioni differenziali.

Costruiamo una successione di funzioni $\mathbf{x}^{(N)}(t)$, continue su $[t_0, t_0 + T]$, che, si può provare, converge a una funzione $\mathbf{x}(t)$ che risolve (10.8). Per questo, dividiamo l'intervallo $[t_0, t_0 + T]$ in N parti uguali mediante i punti

$$t_0\, , \quad t_0 + \delta\, , \quad t_0 + 2\delta\, , \quad \ldots \quad , t_0 + (N - 1)\delta\, , \qquad \text{ove } \delta = T/N .$$

[2] Un possibile valore per T può costruirsi come segue: Sia

$$M = \max_{(t,\mathbf{x})\in R} ||\mathbf{f}(t, \mathbf{x})||\, , \qquad \tilde{M} = \max_i \left\{ \max_{(t,\mathbf{x})\in R} \left|\left| \frac{\partial \mathbf{f}(t, \mathbf{x})}{\partial x_i} \right|\right| \right\} .$$

Si può scegliere

$$T < \min \left\{ a\, ,\ \frac{b}{M}\, ,\ \frac{1}{\tilde{M}} \right\} .$$

Notare che questo valore di T non identifica il più grande intervallo su cui la soluzione esiste.

Inizialmente siamo nel punto $(t_0, \mathbf{x}_0)$. Definiamo la funzione $\mathbf{x}^{(N)}(t)$ sul primo intervallo $[t_0, t_0 + \delta]$ indicandone il grafico: il grafico è il segmento di retta

$$y = \mathbf{x}(t_0) + \mathbf{f}(t_0, \mathbf{x}(t_0))(t - t_0)\,, \qquad t \in [t_0, t_0 + \delta]\,.$$

In questo modo si identifica in particolare $\mathbf{x}^{(N)}(t_0+\delta)$. A partire dal punto $(t_0+\delta, \mathbf{x}^{(N)}(t_0+\delta))$ si ripeta il procedimento: si definisca $\mathbf{x}^{(N)}(t)$ per $t \in [t_0 + \delta, t_0 + 2\delta]$:

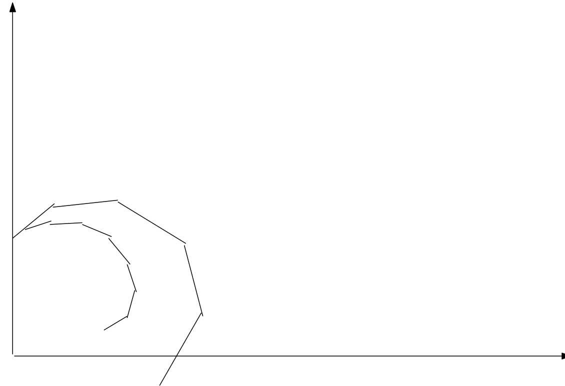
$$\mathbf{x}^{(N)}(t) = \mathbf{x}^{(N)}(t_0 + \delta) + \mathbf{f}(t_0 + \delta, \mathbf{x}^{(N)}(t_0 + \delta))(t - (t_0 + \delta))\,, \qquad t \in [t_0 + \delta, t_0 + 2\delta]\,.$$

Si ripeta quindi questo procedimento fino a coprire tutto l'intervallo $[t_0, t_0 + T]$. In generale avremo, per $k = 0, 1, \ldots, N - 1$,

$$\mathbf{x}^{(N)}(t) = \mathbf{x}^{(N)}(t_0 + k\delta) + \mathbf{f}(t_0 + k\delta, \mathbf{x}^{(N)}(t_0 + k\delta))(t - (t_0 + k\delta))\,, \quad t \in [t_0 + k\delta, t_0 + (k+1)\delta]\,.$$

La figura 10.1 illustra questo procedimento.

Figura 10.1: Spezzata di Eulero



In questo modo, per ogni N, si costruisce una spezzata che si chiama *spezzata* *di Eulero*.

Si può provare che la successione di funzioni $\left\{x^{(N)}(t)\right\}$ converge uniformemente ad una funzione $x(t)$ sull'intervallo $(t_0 - T, t_0 + T)$, purché T sia opportunamente scelto.

Ora, vorremmo sostituire questa funzione nei due membri di (10.8) trovando che i due membri differiscono per un errore che tende a zero per $N \to +\infty$. Però, questo non può farsi perché le funzioni $x^{(N)}(t)$ non sono derivabili; e anche se si potesse fare a poco servirebbe perché la sola convergenza uniforme niente permette di concludere sulla derivata della funzione limite $x(t)$. Viene però in aiuto un'altra idea: integrando i due membri dell'equazione differenziale in (10.8) si vede che $\mathbf{x}(t)$ risolve (10.8) se e solo se essa risolve l'equazione integrale

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{f}(s, \mathbf{x}(s))\,\mathrm{d}s\,. \tag{10.10}$$

E' possibile sostituire la $x^{(N)}(t)$ nei due membri dell'equazione integrale, e stimare lo scarto tra i due membri, provando che esso tende a zero per $N \to +\infty$; e usando la sola convergenza uniforme è ora possibile completare la dimostrazione del teorema.

Omettiamo i dettagli della dimostrazione, contentandoci di avere illustrato queste idee fondamentali. ∎

La dimostrazione del Teorema di Cauchy suggerisce un modo importante di pensare ad un'equazione differenziale, che illustriamo nel caso del sistema

$$x' = f(x, y)\,, \qquad y' = g(x, y)\ :$$

si può pensare al campo vettoriale $f(x, y)\mathbf{i} + g(x, y)\mathbf{j}$ come ad un "campo di velocità" che trasporta un pallino inizialmente nella posizione (x_0, y_0). La traiettoria descritta dal pallino è la soluzione del problema di Cauchy con quel dato iniziale. Da questo punto di vista,

$$\text{il vettore} \quad \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f(x, y) \\ g(x, y) \end{bmatrix} \quad \text{applicato in} \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

rappresenta la velocità che il pallino ha se transita dalla posizione $\mathbf{x}$ all'istante t.

Le isocline

Consideriamo il caso di un sistema autonomo di due equazioni differenziali

$$x' = f(x, y)\,, \qquad y' = g(x, y)\,.$$

La dimostrazione del Teorema di Cauchy suggerisce di considerare il luogo dei punti che vengono incontrati dalle soluzioni dell'equazione differenziale con pendenza assegnata. Escludendo i punti nei quali sia f che g si annullano, le orbite sono curve regolari di parametrizzazione $(x(t), y(t))$. Per fissare le idee, sia $x'(t_0) \neq 0$. La pendenza della tangente nell'istante t_0 è

$$\frac{y'(t_0)}{x'(t_0)} = \frac{g(x(t_0), y(t_0))}{f(x(t_0), y(t_0))}\,.$$

Se invece $x'(t_0) = 0$ allora la tangente è verticale.

Dunque, le orbite hanno tangente verticale nei punti dell'insieme identificato da $f(x, y) = 0$; hanno tangente con pendenza m nei punti dell'insieme identificato da

$$\frac{g(x, y)}{f(x, y)} = m\,.$$

Quest'insieme si chiama l'*isoclina* di pendenza m. L'isoclina di pendenza m ha questa proprietà: le soluzioni escono dai suoi punti con tangenti tutte parallele, con la medesima pendenza m. Se si riescono a disegnare "molte" isocline e quindi a disegnare su di esse segmenti di pendenza m, si può pensare di tracciare una soluzione del problema di Cauchy in questo modo: si parte dal dato (x_0, y_0), che appartiene ad una certa isoclina, corrispondente alla pendenza m_0. Si traccia per (x_0, y_0) un segmento parallelo a $y = m_0 x$ fino ad incontrare la "successiva" isoclina (successiva, tra quelle tracciate) diciamo corrispondente alla pendenza m_1. Dal punto di intersezione si fa uscire un segmento parallelo a $y = m_1 x$ fino ad incontrare l'isoclina successiva, ecc. Si costruisce una spezzata che approssima la soluzione del problema di Cauchy.

10.2.1 Equazioni differenziali lineari a coefficienti costanti

Si è visto, al paragrafo 2.4.4 che l'equazione differenziale lineare a coefficienti costanti

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{f}(t) \tag{10.11}$$

(con A matrice $n \times n$) ha per soluzioni le funzioni

$$\mathbf{x}(t) = e^{At}\mathbf{x_0} + \int_0^t e^{A(t-s)}\mathbf{f}(s)\,\mathrm{d}s\,.$$

Il membro destro di questa formula si chiama *soluzione generale* o anche *integrale generale* della (10.11).

La soluzione generale della (10.11) si ottiene sommando ad una soluzione particolare,

$$\int_0^t e^{A(t-s)}\mathbf{f}(s)\,\mathrm{d}s$$

tutte le funzioni

$$e^{At}\mathbf{x_0}\,, \qquad \mathbf{x_0} \in \mathbf{R}^n\,.$$

Al variare di $\mathbf{x_0}$ queste sono tutte le soluzioni dell'*equazione lineare omogenea associata*, ossia dell'equazione che si ottiene dalla (10.11) ponendo $\mathbf{f}(t) = 0$.

Dunque, il calcolo dell'integrale generale si riduce al calcolo dell'esponenziale di matrici

$$e^{At}\,.$$

Il calcolo di quest'esponenziale è tutt'altro che semplice in generale e ci limitiamo a studiare il caso dei "sistemi piani", ossia il caso in cui $\mathbf{x} \in \mathbf{R}^2$.

Consideriamo prima di tutto l'equazione lineare omogenea, ossia assumiamo $\mathbf{f}(t) = 0$. Dobbiamo ricordare alcune proprietà delle matrici, che richiamiamo esclusivamente nel caso di matrici 2×2. Proprietà analoghe valgono in generale, ma la casistica diventa sempre più complessa all'aumentare delle dimensioni della matrice.

Ogni matrice ammette almeno un *autovalore* ossia almeno un numero λ tale che

$$\det[A - \lambda I] = 0\,.$$

Questo discende dal teorema fondamentale dell'algebra, perché $\det[A - \lambda I]$ è un polinomio non costante, e quindi ammette almeno uno zero che però può essere un numero complesso, anche se gli elementi della matrice sono reali.

Il polinomio $\det[A - \lambda I]$ si chiama *polinomio caratteristico* della matrice A.

In corrispondenza di ciascun autovalore, si trovano soluzioni $\mathbf{v}$ non nulle dell'equazione lineare

$$A\mathbf{v} = \lambda\mathbf{v}\,.$$

Questi vettori non nulli si chiamano gli *autovettori* relativi all'autovalore λ.

Naturalmente, se $\mathbf{v}$ è un autovettore relativo a λ, anche i suoi multipli lo sono.

Dobbiamo osservare che:

- se λ è un numero complesso, anche i suoi autovettori sono autovettori a elementi complessi; ossia, implicitamente anche se gli elementi della matrice A sono reali, ci riconduciamo a lavorare con vettori ad elementi complessi;
- autovettori relativi ad autovalori diversi sono linearmente indipendenti.

Consideriamo la funzione

$$\mathbf{x}(t) = e^{At}\mathbf{v}\,.$$

Allora,

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{At}\mathbf{v} = Ae^{At}\mathbf{v}$$

è una soluzione della (10.11) (anche se $\mathbf{v}$ non è autovettore di A). Se

$$A\mathbf{v} = \lambda\mathbf{v}\,,$$

si ha

$$e^{At}\mathbf{v} = \left[\sum_{n=0}^{+\infty}\frac{A^n}{n!}\right]\mathbf{v} = \left[\sum_{n=0}^{+\infty}\frac{A^n\mathbf{v}}{n!}\right] = \sum_{n=0}^{+\infty}\frac{\lambda^n\mathbf{v}}{n!}$$

ossia

$$e^{At}\mathbf{v} = e^{\lambda t}\mathbf{v}\,.$$

Dunque, noto un autovalore λ e un suo autovettore $\mathbf{v}$, è facile scrivere alcune soluzioni dell'equazione lineare omogenea: sono le funzioni

$$\alpha e^{\lambda t}\mathbf{v}$$

con α coefficiente qualsiasi (in generale, coefficiente complesso).

Esista ora un secondo autovalore μ e sia $\mathbf{w}$ un suo autovettore. Allora, anche le funzioni

$$\beta e^{\mu t}\mathbf{w}$$

risolvono l'equazione ldifferenziale lineare omogenea

$$\mathbf{x}' = A\mathbf{x}\,; \tag{10.12}$$

e quindi anche le funzioni

$$\alpha e^{\lambda t}\mathbf{v} + \beta e^{\mu t}\mathbf{w} \tag{10.13}$$

la risolvono.

Quanto detto fin'ora vale per sistemi di qualunque dimensione. Supponiamo ora che il sistema sia di dimensione 2.

Dato che i vettori $\mathbf{v}$ e $\mathbf{w}$ sono linearmente indipendenti, ogni vettore a due dimensioni $\mathbf{x}$ si può rappresentare in modo unico come

$$\mathbf{x} = \alpha\mathbf{v} + \beta\mathbf{w}\,;$$

Ossia, al variare dei coefficienti α e β, le funzioni (10.13) permettono di esprimere tutte le soluzioni di (10.12). Dunque, la (10.13) è la soluzione generale di (10.12).

Si noti che il fatto importante è l'esistenza di due autovettori $\mathbf{v}$ e $\mathbf{w}$ linearmente indipendenti: le considerazioni precedenti possono ripetersi anche se i due autovettori $\mathbf{v}$ e $\mathbf{w}$ corrispondono ad un medesimo autovalore, ossia se $\lambda = \mu$.

Sfortunatamente, non è vero che ogni matrice 2×2 debba avere due autovettori linearmente indipendente, come mostra l'esempio seguente:

Esempio 271 Gli autovettori della matrice

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

sono tutti i multipli di

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} :$$

Questa matrice non ha due autovettori linearmente indipendenti. ∎

In tal caso, si procede come segue: si identifica (l'unico) autovettore λ ed un suo autovettore $\mathbf{v}$. Ricordiamo che per definizione $\mathbf{v} \neq 0$. Quindi si considera l'equazione

$$(A - \lambda I)\,\mathbf{w} = \mathbf{v}\,.$$

E' un fatto che, se A, matrice 2×2, non ha due autovettori linearmente indipendenti, quest'equazione ammette soluzione $\mathbf{w}$, e inoltre $\mathbf{w}$ è linearmente indipendente da $\mathbf{v}$.

Identificati $\mathbf{v}$ e $\mathbf{w}$, si costruiscono due polinomi, a valori vettori, uno di grado 0 ed uno di grado 1:

$$\mathbf{p_0} = \mathbf{v}\,, \qquad \mathbf{p_1} = \mathbf{v}t + \mathbf{w}\,.$$

Nel caso che stiamo ora considerando, si verifica facilmente per sostituzione che la soluzione generale di (10.12) è

$$e^{\lambda t}\left[\alpha \mathbf{p_0} + \beta \mathbf{p_1}(t)\right]\,. \tag{10.14}$$

Abbiamo quindi scritto la soluzione generale di (10.12) in ciascuno dei casi che possono presentarsi: in tutti i casi la soluzione generale è combinazione lineare di due soluzioni fondamentali. Delle soluzioni

$$e^{\lambda t}\mathbf{p_0}\,, \qquad e^{\lambda t}\mathbf{p_1}(t)$$

delle soluzioni

$$e^{\tau t}\mathbf{v}\,, \qquad e^{\mu t}\mathbf{w}$$

(con $\lambda = \mu$ oppure $\lambda \neq \mu$) quando la matrice A ammette due autovettori linearmente indipendenti. Queste coppie di soluzioni di (10.12) si chiamano un *sistema fondamentale* di soluzioni.

Rimane da chiarire un punto: se la matrice A è reale ma ammette due autovalori complessi e coniugati, le soluzioni scritte sopra prendono valori complessi. Vorremmo trovare un diverso sistema fondamentale di soluzioni, che però prendano valori reali.

Il caso degli autovettori complessi e coniugati Le considerazioni precedenti valgono per ogni matrice A sia che i suoi autovalori siano reali sia che siano complessi. Se però la matrice 2×2 ha elementi reali ed un autovalore λ complesso, allora anche $\bar{\lambda}$ è autovalore; e quindi siamo automaticamente nel caso in cui la matrice ha due diversi autovalori e quindi due autovettori linearmente indipendenti: la soluzione generale di (10.12) è data da (10.13).

Siano

$$\lambda_{\pm} = \xi \pm i\omega$$

gli autovalori e sia $\mathbf{v}_+ = \mathbf{v_1} + i\mathbf{v}_2$ un autovettore di λ_+. In quest'espressione, $\mathbf{v_1}$ e $\mathbf{v_2}$ sono vettori reali, che però non sono essi stessi autovettori di A. E' un fatto che

$$\bar{\mathbf{v}}_+ = \mathbf{v_1} - i\mathbf{v}_2$$

è un autovettore di λ_-. Usando la (10.13) si vede quindi che un sistema fondamentale di soluzioni è

$$\begin{array}{l} e^{\xi t}\left[e^{-i\omega t}\left(\mathbf{v_1} + i\mathbf{v_2}\right)\right] \\ = e^{\xi t}\left[\left(\mathbf{v_1}\cos\omega t - \mathbf{v_2}\sin\omega t\right) + i\left(\mathbf{v_2}\cos\omega t + \mathbf{v_1}\sin\omega t\right)\right] \\ e^{\xi t}\left[e^{-i\omega t}\left(\mathbf{v_1} - i\mathbf{v_2}\right)\right] \\ = e^{\xi t}\left[\left(\mathbf{v_1}\cos\omega t - \mathbf{v_2}\sin\omega t\right) - i\left(\mathbf{v_2}\cos\omega t + \mathbf{v_1}\sin\omega t\right)\right] . \end{array} \tag{10.15}$$

Queste soluzioni prendono valori complessi. Sommandole e sottraendole (e dividendo rispettivamente per 2 e per $2i$) si trovano due soluzioni a valori reali:

$$e^{\xi t}\left[\mathbf{v_1}\cos\omega t - \mathbf{v_2}\sin\omega t\right] , \qquad e^{\xi t}\left[\mathbf{v_1}\cos\omega t + \mathbf{v_2}\sin\omega t\right] .$$

Questo anche è un sistema fondamentale di soluzioni, a valori reali. Infatti, combinandole linearmente si ritrovano le (10.15) e quindi ogni altra soluzione di (10.12).

10.2.2 Equazioni completi ed equazioni di ordine superiore

Limitandoci ancora al caso di sistemi di due equazioni, mostriamo un metodo per risolvere l'equazione completa. Abbiamo bisogno di un risultato di algebra lineare, che è questo: scriviamo esplicitamente il sistema in forma

$$\left\{ \begin{array}{lcl} x' & = & ax + by + f_1 \\ y' & = & cx + dy + f_2 \end{array} \right. \tag{10.16}$$

e consideriamo la matrice

$$A = \left[\begin{array}{cc} a & b \\ c & d \end{array} \right] .$$

Mostriamo:

Teorema 272 *Esiste una matrice P (generalmente ad elementi complessi) tale che* $\det P \neq 0$ *e inoltre tale che*

$$PAP^{-1} = P \left[\begin{array}{cc} a & b \\ c & d \end{array} \right] P^{-1} = \left[\begin{array}{cc} \alpha & \beta \\ 0 & \delta \end{array} \right] .$$

Dim. Bisogna ricordare che ogni matrice ammette almeno un autovalore ed un corrispondente autovettore, che in generale sono complessi. Dunque si trovano $\mathbf{v}$ e α tali che

$$A\mathbf{v} = \alpha\mathbf{v} .$$

Sia $\mathbf{w}$ un qualsiasi vettore indipendente da $\mathbf{v}$ e consideriamo la matrice

$$P^{-1} = \left[\begin{array}{cc} \mathbf{v} & \mathbf{w} \end{array} \right] .$$

Questa matrice trasforma gli elementi della base canonica ordinatamente in $\mathbf{v}$ e $\mathbf{w}$. La sua inversa quindi trasforma ordinatamente $\mathbf{v}$ e $\mathbf{w}$ nella base canonica. E':

$$AP^{-1} = \begin{bmatrix} \alpha\mathbf{v} & \mathbf{z} \end{bmatrix} \qquad \mathbf{z} = A\mathbf{w} = \beta\mathbf{v} + \delta\mathbf{w}$$

e quindi

$$PAP^{-1} = \begin{bmatrix} \alpha & \beta \\ 0 & \delta \end{bmatrix} . \quad \blacksquare$$

Sia ora

$$\mathbf{g}(t) = \begin{bmatrix} g_1(t) \\ g_2(t) \end{bmatrix} = P\mathbf{f}(t) = P \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}$$

Applicando la trasformazione P al sistema, questo si riduce a

$$\xi' = \alpha\xi + \beta\eta + g_1(t) , \qquad \eta' = \delta\eta + g_2(t) .$$

La seconda equazione è un'equazione nella sola incognita $\eta(t)$, che si sa risolvere; nota $\eta(t)$, questa funzione si sostituisce nella prima equazione, ottenendo un'equazione nella sola incognita ξ, che ancora si sa risolvere.

Calcolate le funzioni $\xi(t)$ ed $\eta(t)$, la soluzione $\mathbf{x}(t)$ si ottiene come

$$\mathbf{x}(t) = P^{-1} \begin{bmatrix} \xi(t) \\ \eta(t) \end{bmatrix} .$$

Quando il termine noto ha forma "particolare", dal corso di Analisi Matematica 1, si sa che conviene ricercare una soluzione di forma "particolare". Consideriamo quest'esempio:

$$x' = \alpha x + \beta y + e^{\gamma t} , \qquad y' = \delta y + e^{\sigma t} .$$

La risoluzione di questo sistema può farsi semplicemente con le tecniche viste nel corso di Analisi Matematica 1 e quindi non entriamo nei dettagli del calcolo. Notiamo però che se $\delta \neq \sigma$ allora le soluzioni della seconda equazione hanno forma

$$ae^{\delta t} + be^{\sigma t} .$$

Se però $\delta = \sigma$ allora le soluzioni sono

$$ae^{\delta t} + bte^{\delta t} .$$

Inserite queste nella prima equazione, la soluzione della prima equazione ha forma rispettivamente

$$ce^{\alpha t} + ae^{\delta t} + be^{\sigma t} \quad \text{oppure} \quad ce^{\alpha t} + ae^{\delta t} + bte^{\delta t}$$

se α è diverso dagli esponenti che si sono incontrati risolvendo la seconda equazione; altrimenti compariranno fattori t^n con $n > 1$. Si esaminino i vari casi e si ritrovi che i possibili valori di n sono 0, 1 e 2.

E ora consideramo il caso di un sistema omogeneo di tre equazioni differenziali lineari. Anche in questo caso esistono trasformazioni di coordinate che "triangolarizzano" il sistema, ossia lo riducono a forma

$$\begin{cases} x' &= \alpha x + \beta y + \chi z \\ y' &= \delta y + \sigma z \\ z' &= \gamma z \end{cases}$$

Anche questo sistema si risolve a partire dall'ultima equazione e, risoltala, si vede che $z(t) = ae^{\gamma t}$. Sostituita queste funzione nelle prime due equazioni, ci si riconduce al caso di due equazioni lineari con termine affine visto sopra. E quindi nell'espressione delle prime due componenti compariranno termini della forma $t^n e^{\gamma t}$ con $n = 0$, $n = 1$ oppure $n = 2$.

10.2.3 Il comportamento in futuro delle soluzioni

Spesso interessa conoscere il comportamento delle soluzioni di un sistema di equazioni differenziali per $t \to +\infty$. Quando si fa questo studio, si dice che si studia il "comportamento in futuro" delle soluzioni. Facciamo questo studio nel caso dei sistemi lineari omogenei di due equazioni differenziali. In tal caso, le soluzioni sono funzioni $\mathbf{x}(t)$ con $\mathbf{x}(t)$ vettore a due componenti $x(t)$ ed $y(t)$. Dunque le soluzioni rappresentano curve parametriche piane. Per questo, i sistemi due due equazioni differenziali si chiamano anche *sistemi piani* ed il piano (x, y) si chiama *piano delle fasi*.

Vogliamo descrive le soluzioni di un sistema lineare omogeneo piano sul piano delle fasi e vedere come il comportamento dipenda dagli autovalori della matrice A e, se c'è un solo autovalore, anche dal fatto che esistano due autovettori linearmente indipendenti, o uno solo.

La soluzione stazionaria $\mathbf{x}(t) = \mathbf{0}$ esiste sempre. Noi considereremo solamente il caso in cui

$$\ker A = \mathbf{0}$$

ossia in cui il sistema differenziale ammette un'unica soluzione stazionaria, e questa è $\mathbf{x}(t) = \mathbf{0}$. Interessa conoscere il comportamente delle altre soluzioni rispetto alla soluzione stazionaria. Questo comportamento viene descritto introducendo opportune definizioni[3]. Nel caso lineare però non abbiamo bisogno dei dettagli di queste definizioni. Diciamo per ora che interessa sapere se tutte le soluzioni rimangono limitate e se tutte le soluzioni tendono a zero per $t \to +\infty$. Naturalmente tutte le soluzioni rimangono limitate se le due soluzioni che costituiscono un sistema fondamentale di soluzioni sono ambedue limitate; tutte le soluzioni tendono a zero per $t \to +\infty$ se ciò accade per ambedue gli elementi di un sistema fondamentale di soluzioni.

Ricapitoliamo i casi che si possono presentare: casi:

1) la matrice A ha due autovettori indipendenti $\mathbf{v}_1$ e $\mathbf{v}_2$, corrispondenti rispettivamente agli autovalori reali λ_1 e λ_2 (non necessariamente distinti). Allora un sistema fondamentale di soluzioni è dato da

$$\mathbf{x}_1(t) = e^{\lambda_1 t}\mathbf{v}_1, \quad \mathbf{x}_2(t) = e^{\lambda_2 t}\mathbf{v}_2;$$

2) la matrice A ha due autovettori complessi coniugati $\mathbf{v} \pm i\mathbf{w}$ e corrispondenti rispettivamente agli autovalori complessi coniugati $\alpha \pm i\beta$. Un sistema fondamentale di soluzioni è dato da

$$\mathbf{x}_1(t) = e^{\alpha t}(\cos(\beta t)\mathbf{v} - \sin(\beta t)\mathbf{w}), \quad \mathbf{x}_2(t) = e^{\alpha t}(\cos(\beta t)\mathbf{w} + \sin(\beta t)\mathbf{v});$$

3) la matrice A ha un solo autovalore λ ma non vi sono due autovettori indipendenti. Allora, detto $\mathbf{v}_1$ un autovettore, si risolve il sistema[4] $(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1$ e un sistema fondamentale di soluzioni è dato da

$$\mathbf{x}_1(t) = e^{\lambda t}\mathbf{v}_1, \quad \mathbf{x}_2(t) = e^{\lambda t}(t\mathbf{v}_1 + \mathbf{v}_2).$$

Esaminando separatamente questi tre casi, si vede che valgono i risultati elencati nella tabella seguente:

[3] di "stabilità", che vedremo piu' avanti nel caso non lineare.

[4] si tratta di un sistema la cui matrice dei coefficienti è degenere, ma si dimostra che è compatibile

Tabella 10.1: Comportamento asintotico delle soluzioni dei sistemi lineari piani

Due autovettori linearm. indipendenti	Soluzioni limitate in futuro se e solo se ambedue gli autovalori hanno parte reale non positiva
Due autovettori linearm. indipendenti	Le soluzioni tendono a zero per $t \to +\infty$ se e solo se ambedue gli autovalori hanno parte reale negativa
Autovettori tutti linearmente dipendenti (e quindi un solo autovalore δ)	le soluzioni sono limitate in futuro se e solo se δ è negativo (tutte le soluzioni tendono a zero per $t \to +\infty$).

I casi che possono presentarsi sono ben rappresentati negli esempi seguenti.

Esempio 273

- il caso di due autovalori reali distinti e di segno concorde. In questo caso le soluzioni del sistema diagonalizzato hanno forma

$$x(t) = e^{\lambda t} x_0\,, \qquad y(t) = e^{\mu t} y_0\,.$$

Le orbite sono le curve

$$y^\lambda = c x^\mu$$

percorse verso l'origine se gli autovalori sono negativi, allontanandosi dall'origine altrimenti.

Se $\lambda = \mu$ le orbite sono rette.

Questo caso è illustrato dalla figura 10.2 a sinistra. Questa configurazione di orbite si chiama *nodo.*

- I due autovalori sono reali, di segno opposto. In questo caso le soluzioni sono della forma

$$x(t) = e^{\lambda t} x_0\,, \qquad y(t) = e^{-\mu t} y_0$$

con, per esempio, $\lambda > 0$ e $\mu > 0$. Nel caso $\lambda = \mu$ le orbite sono le iperboli

$$xy = \text{cost.}$$

e per questa ragione il punto di equilibrio $(0, 0)$ si dice di tipo *iperbolico,* o di *sella.* La figura (10.2) a destra illustra questo caso.

- I due autovalori sono puramente immaginari, e coniugati. In questo caso le soluzioni hanno forma

$$x(t) = A\cos(\omega t + \phi)\,, \qquad y(t) = A\sin(\omega t + \phi)\,.$$

Le orbite sono le circonferenze

$$x^2 + y^2 = A^2\,.$$

Il caso è illustrato nella figura 10.3 a sinistra, e il punto di equilibrio si chiama ora *centro.*

Figura 10.2: Nodo e punto di sella

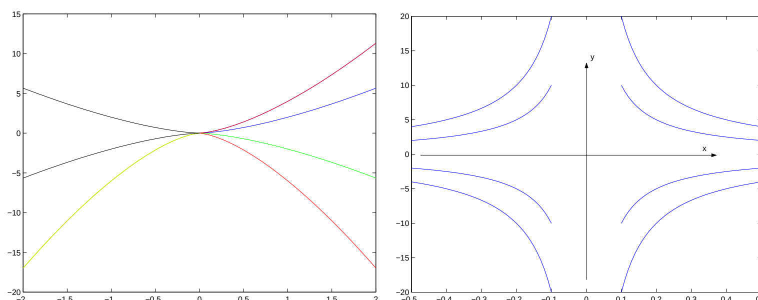
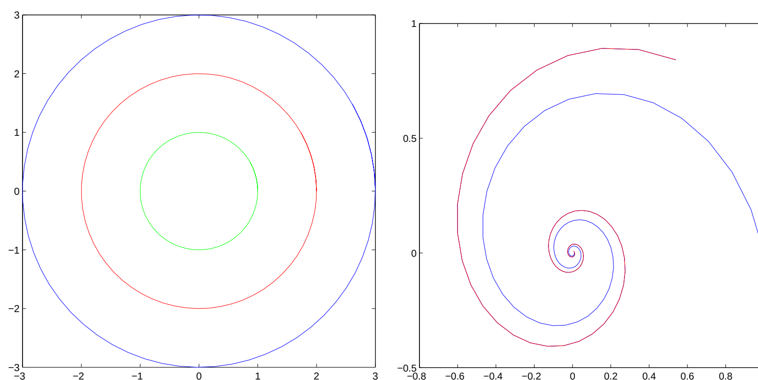


Figura 10.3: Centro e fuoco



- I due valori sono numeri complessi e coniugati, ma non puramente immaginari. In questo caso le soluzioni hanno forma

$$x(t) = Ae^{at}\cos(\omega t + \phi)\,, \qquad y(t) = Ae^{at}\sin(\omega t + \phi)\,.$$

  In coordinate polari sono descritte da

$$r(t) = Ae^{at}\,, \qquad \theta(t) = \omega t + \phi$$

  e sono quindi spirali che si avvolgono verso l'origine se $a < 0$, altrimenti si allontanano dall'origine. Il punto d'equilibrio $(0, 0)$ è ora detto *fuoco*. Questo caso è illustrato nella figura 10.3, a destra.

  Il numero a è la parte reale comune ai due autovalori.

10.3 La stabilità

Consideriamo ora il sistema non lineare autonomo[5]

$$\mathbf{x}' = \mathbf{f}(\mathbf{x})\,. \qquad (10.17)$$

Per semplicità supponiamo che tutte le soluzioni di (10.17) siano definite su $[0, +\infty)$.[6]

Ricordiamo che i punti di equilibrio sono quei punti $\mathbf{x_0}$ tali che $\mathbf{f}(\mathbf{x_0}) = \mathbf{0}$. I punti di equilibrio corrispondono a soluzioni costanti e spesso queste sono le "più importanti" tra le soluzioni dell'equazione differenziale. Per questa ragione, interessa sapere come si comportano soluzioni il cui dato iniziale poco differisce da un punto di equilibrio; in particolare interessa sapere se per $t \to +\infty$ tali soluzioni tendono o meno al punto di equilibrio stesso.

Introduciamo le definizioni seguenti.

Definitione 9 Diciamo che tutte le soluzioni di (10.17) sono *limitate in futuro* quando per ogni $\mathbf{x}_0$ esiste $M_{\mathbf{x}_0}$ tale che

$$||\mathbf{x}(t; \mathbf{x}_0)|| < M_{\mathbf{x}_0} \qquad \forall t > 0\,.$$

In modo analogo si definisce la *limitatezza in passato.*.

Si dice che le soluzioni di (10.17) sono *limitate* quando sono limitate sia in futuro che in passato. ∎

Esempio 274 Consideriamo l'equazione scalare

$$x' = ax\,,$$

le cui soluzioni sono $x(t) = ce^{at}$. Tutte le soluzioni sono limitate in futuro ma non in passato se $a < 0$; sono limitate in passato ma non in futuro se $a > 0$; sono limitate se $a = 0$. ∎

Presentiamo ora le definizioni di stabilità, con riferimento ad un punto di equilibrio $\mathbf{y}_0$.

Definitione 10 *Diciamo che il punto di equilibrio $\mathbf{y}_0$ è*

- *stabile* *quando per ogni $\epsilon > 0$ esiste $\sigma > 0$ tale che se $||\mathbf{x}_0 - \mathbf{y}_0|| < \sigma$ allora $||\mathbf{x}(t; \mathbf{x}_0) - \mathbf{y}_0|| < \epsilon$ per ogni $t > 0$ (si noti: $\sigma \le \epsilon$).*
- *attrattivo* *se esiste $\sigma > 0$ tale che se $||\mathbf{x}_0 - \mathbf{y}_0|| < \sigma$ allora $\lim_{t\to+\infty} ||\mathbf{x}(t; \mathbf{x}_0)|| = \mathbf{y}_0$. L'insieme dei vettori $\mathbf{x}_0$ tali che $\lim_{t\to+\infty} ||\mathbf{x}(t; \mathbf{x}_0)|| = \mathbf{y}_0$ si chiama il* *bacino d'attrazione* *di $\mathbf{y}_0$.*
- *asintoticamente stabile* *se è sia stabile che attrattivo.*

Osservazione 275 E' bene notare che l'attrattività non implica la stabilità. ∎

---

[5]le definizioni di questo paragrafo possono adattarsi anche al caso dei sistemi non autonomi. Tal caso è più delicato e non lo consideriamo.

[6]non è restrittivo assumere che l'istante iniziale t_0 sia 0 perché il sistema è tempo invariante.

Si potrebbe pensare che se $\mathbf{y}_0$ è un punto di equilibrio stabile allora si possa sempre scegliere $\delta = \epsilon$. L'esempio seguente mostra che questa congettura è falsa:

Esempio 276 Si consideri il sistema

$$x' = 2y\,, \qquad y' = -x\,.$$

L'unico punto di equilibrio è $\mathbf{y}_0 = (0,0)$.

Derivando la prima equazione si vede che

$$x'' = -2x$$

e questa è l'equazione del moto armonico. Dunque, le soluzioni del sistema sono

$$x(t) = A\cos\sqrt{2}t + \phi\,, \qquad y(t) = -\frac{A}{\sqrt{2}}\sin(\sqrt{2}t + \phi)\,.$$

Dunque le soluzioni descrivono l'ellisse

$$x^2 + 2y^2 = 1\,.$$

Sono tutte limitate e inoltre se si assegna ϵ si può trovare un δ tale che

$$||(x_0, y_0)|| < \delta \implies ||(x(t), y(t))|| < \epsilon\,.$$

Però, δ va scelto strettamente minore di ϵ, si veda la figura 10.4. ∎

Figura 10.4: Nella definizione di stabilità in generale si deve scegliere $\delta < \epsilon$

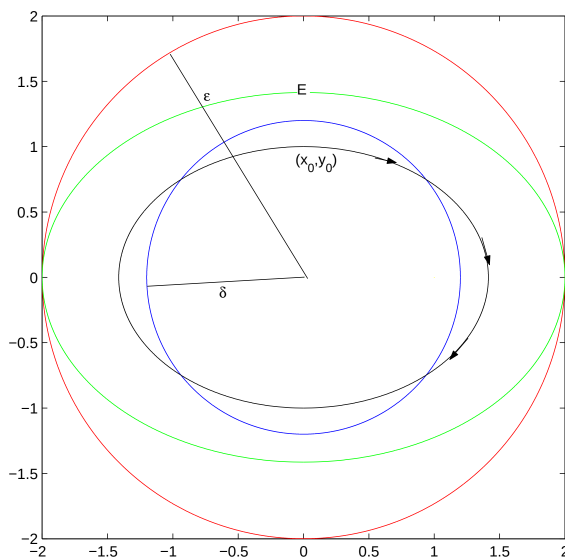


Tabella 10.2: La stabilità dei sistemi lineari piani

Due autovettori linearm. indipendenti	Soluzione nulla stabile se e solo se ambedue gli autovalori hanno parte reale non positiva
Due autovettori linearm. indipendenti	Soluzione nulla asintoticamente stabile se e solo se ambedue gli autovalori hanno parte reale negativa
Autovettori tutti linearmente dipendenti (e quindi un solo autovalore δ)	soluzione nulla asintoticamente stabile se e solo se l'unico autovalore è negativo

Quest'esempio verrà nuovamente esaminato in seguito.

Osserviamo che se il punto di equilibrio $\mathbf{y}_0$ è stabile allora le soluzioni (con dato iniziale sufficientemente vicino a $\mathbf{y}_0$) sono limitate in futuro e addirittura in modo uniforme: se $||\mathbf{x}_0 - \mathbf{y}_0|| < \delta$ allora $||\mathbf{x}(t;\mathbf{x}_0)|| < \epsilon + ||\mathbf{y}_0|| = M_{\mathbf{x}_0}$. In questo caso $M_{\mathbf{x}_0} = \epsilon + ||\mathbf{y}_0||$ è indipendente da $\mathbf{x}_0$ (se $||\mathbf{x}_0 - \mathbf{y}_0|| < \delta$). Il viceversa non vale:

Esempio 277 Consideriamo il sistema descritto in coordinate polari da

$$r' = r(1-r)\,, \qquad \theta' = 1\,.$$

$r(t) = 1$ è una soluzione costante e, procedendo per separazione di variabili,

$$\log \frac{r(t)}{|1 - r(t)|} = t + c$$

(si ricordi che $r \geq 0$). Dunque, le soluzioni sono

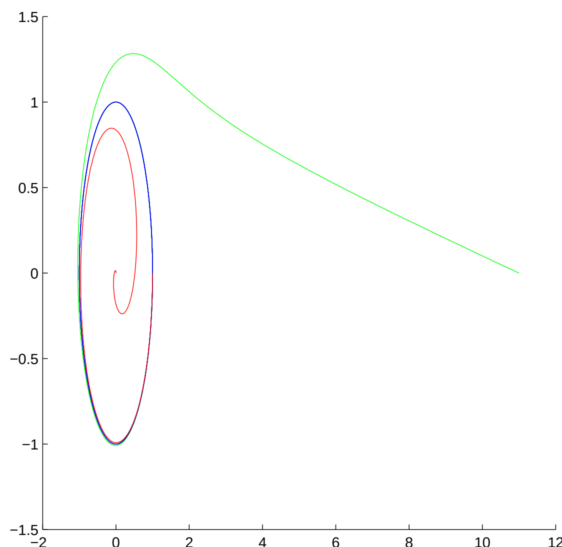
$$\theta(t) = t + h\,, \qquad r(t) = \begin{cases} \frac{ke^t}{1+ke^t} & \text{con } k > 0 \text{ se } 0 \leq r(0) < 1 \\ 1 & \text{se } r(0) = 1 \\ \frac{ke^t}{ke^t - 1} & \text{con } k > 1 \text{ se } r(0) > 1. \end{cases}$$

Le soluzioni descrivono delle spirali che si avvolgono intorno alla circonferenza di raggio 1. Tutte le soluzioni sono limitate ma scegliendo un intorno dell'origine di raggio minore di 1, non si riesce a rimanere in quest'intorno scegliendo di partire "vicini" all'origine. Si veda la figura 10.5. ∎

Concludiamo notando che i risultati della tabella 10.1, che si riferisce ai sistemi lineari piani autonomi, possono riformularsi con la terminologia della stabilità. Si ottiene così la tabella 10.2, nella quale $\mathbf{y}_0 = 0$.

Si pone ora il problema di sapere quando un punto di equilibrio di un sistema piano è stabile oppure asintoticamente stabile. Questo è generalmente un problema piuttosto difficile. I due test più semplici che possono usarsi sono il Teorema di Lagrange, per la stabilità, e il teorema sulla stabilità "in prima approssimazione", per la stabilità asintotica. Per chiarire il significato del Teorema di Lagrange, conviene introdurre la nozione di integrale primo, che ha un'importanza fondamentale in tutte le applicazioni della matematica.

Figura 10.5: Il caso dell'esempio 277: la limitatezza delle soluzioni non implica la stabilità



10.4 Sistemi piani ed integrali primi

Il concetto di integrale primo non è limitato ai sistemi piani, ma è in questo caso particolare che noi lo studieremo. Consideriamo quindi il sistema piano

$$x' = f(x, y)\,, \qquad y' = g(x, y)\,. \tag{10.18}$$

Si chiama *integrale primo* una funzione $V(x, y)$ di classe C^1 tale che:

- gli zeri di $\nabla V(x, y)$ sono isolati;
- ogni orbita di (10.18) è contenuta in una curva di livello di $V(x, y)$.

Integrali primi possono non esistere, ma se esistono non sono mai unici: se $V(x, y)$ è un integrale primo, anche $cV(x, y)$ e $V(x, y) + c$ (con c costante) lo sono.

E' un fatto che molto spesso quando un integrale primo si può trovare allora questo dà informazioni importanti sul comportamento delle soluzioni dell'equazione differenziale. Per esempio:

- se si riesce a trovare un integrale primo allora le orbite del sistema (10.18) si possono (almeno localmente) calcolare risolvendo l'equazione (non differenziale) $V(x, y) = c$. Grazie al Teorema della funzione implicita, ciò è lecito nell'intorno dei punti nei quali il gradiente di $V(x, y)$ non si annulla.
- Se le curve di livello di un integrale primo sono tutte limitate, allora le soluzioni dell'equazione rimangono limitate al trascorrere del tempo.

Gli integrali primi hanno una relazione importante con le proprietà dei campi vettoriali. Provando il Teorema di Cauchy, abbiamo visto che al sistema (10.18) conviene associare il campo vettoriale

$$\mathbf{F}(x,y) = f(x,y)\mathbf{i} + g(x,y)\mathbf{j}\,.$$

Associamogli anche il campo vettoriale

$$\mathbf{G}(x,y) = -g(x,y)\mathbf{i} + f(x,y)\mathbf{j}$$

ortogonale a $\mathbf{F}(x,y)$ in ogni punto.

Supponiamo che $\mathbf{G}(x,y)$ ammetta potenziale $V(x,y)$. Allora, lungo le soluzioni di (10.18) si ha:

$$\frac{\mathrm{d}}{\mathrm{d}t}V(x(t),y(t)) = V_x x' + V_y y' = -gf + fg = 0\,.$$

Ossia, V è un integrale primo di (10.18). Condizione necessaria per l'esistenza del potenziale $V(x,y)$ del campo $\mathbf{G}(x,y)$ è che

$$-g_y = f_x$$

Dunque:

Teorema 278 *Supponiamo che il dominio del sistema piano sia una regione di Jordan. Se*

$$\operatorname{div}\mathbf{F}(x,y) = 0$$

allora il sistema piano ammette integrali primi.

Ovviamente, esistono sistemi che non ammettono integrali primi. Esistono però anche sistemi che ammettono integrali primi e che non soddisfano alle condizioni del Teorema 278, che dà una condizione solo sufficiente per l'esistenza di integrali primi:

Esempio 279 Si consideri il sistema piano

$$t' = 1\,, \qquad x' = x\,.$$

Chiaramente, la condizione del Teorema 278 non è soddisfatta, ma integrali primi esistono:

$$V(t,x) = xe^{-t}$$

è infatti un integrale primo. ∎

Ciò nonostante, il caso descritto nel Teorema 278 è particolarmente importante per le applicazioni. Indicando con $H(x,y)$ un integrale primo, si ha $\nabla H = \begin{bmatrix} -g & f \end{bmatrix}$, ossia

$$f = H_y\,, \qquad g = -H_x$$

e quindi il sistema piano (10.18) si rappresenta come

$$x' = H_y\,, \qquad y' = -H_x\,.$$

Sistemi di questa forma si chiamano *sistemi hamiltoniani* e l'integrale primo $H(x,y)$ si chiama l'*hamiltoniana* del sistema.

Un caso molto importante nel quale è facile trovare un integrale primo si ha quando si studia

$$x'' = -\phi(x) \qquad \text{ossia} \qquad \begin{cases} x' & = & y \\ y' & = & -\phi(x) \end{cases} \tag{10.19}$$

Si noti che ϕ dipende dalla sola posizione x e non dalla velocità y. Assumiamo $\phi(x) \in C^1(\mathbf{R})$. In questo caso

$$\mathbf{F}(x, y) = y\mathbf{i} - \phi(x)\mathbf{j}$$

ha divergenza nulla e quindi il sistema è hamiltoniano. Calcolando il potenziale di

$$\mathbf{G}(x, y) = \phi(x)\mathbf{i} + y\mathbf{j}$$

si trova l'integrale primo

$$V(x, y) = \frac{1}{2}y^2 + \Phi(x)\,, \qquad \Phi(x) = \int_0^x \phi(s)\,\mathrm{d}s\,.$$

Nelle applicazioni fisiche, il sistema si incontra generalmente scritto nella forma

$$mx'' = -\phi(x)$$

ove m è la massa. In tal caso

$$\mathbf{F}(x, y) = y\mathbf{i} - \frac{1}{m}\phi(x)\mathbf{j}\,, \qquad \mathbf{G}(x, y) = \frac{1}{m}\phi(x)\mathbf{i} + y\mathbf{j}$$

ed un integrale primo è

$$V(x, y) = \frac{1}{m}\Phi(x) + \frac{1}{2}y^2\,.$$

Si è detto che i multipli di un integrale primo sono ancora integrali primi. Il particolare multiplo

$$E(x, y) = \frac{1}{2}my^2 + \Phi(x)$$

si chiama *energia totale* del sistema, mentre si chiamano rispettivamente *energia cinetica* ed *energia potenziale* le due funzioni $my^2/2$ e $\Phi(x)$.

I sistemi della forma (10.19) si dicono *conservativi*, poiché l'energia totale (che è un integrale primo) rimane costante lungo ogni orbita (in generale ha valore diverso su orbite diverse):

Teorema 280 *L'energia totale "si conserva" (ossia rimane costante) sulle traiettorie di un sistema conservativo.*

Esempio 281 Consideriamo l'equazione

$$mx'' = -\frac{l}{g}\sin x$$

che rappresenta le oscillazioni di un pendolo di lunghezza costante.[7] La sua energia totale è

$$E(x, y) = \frac{1}{2}my^2 + \frac{l}{g}\int_0^x \sin s\,\mathrm{d}s = \frac{1}{2}my^2 - \frac{l}{g}\left[\cos x - 1\right]\,.$$

[7]si noti che se le oscillazioni sono "piccole" si può approssimare $\sin x \sim x$ ed ottenere l'equazione del moto armonico, $x'' = -\omega^2 x$ con $\omega^2 = l/mg$.

I potenziali sono sempre definiti a meno di una costante additiva, ed usa scegliere come energia totale la funzione

$$E(x,y) = \frac{1}{2}my^2 - \frac{l}{g}\cos x \,. \quad \blacksquare$$

10.4.1 Integrali primi e stabilità

Se si possono studiare le curve di livello di un integrale primo è possibile ottenere informazioni sulla stabilità di un punto di equilibrio. Mostriamo questo su un esempio.

Esempio 282 Si consideri il sistema

$$x' = 2y\,, \qquad y' = -x\,.$$

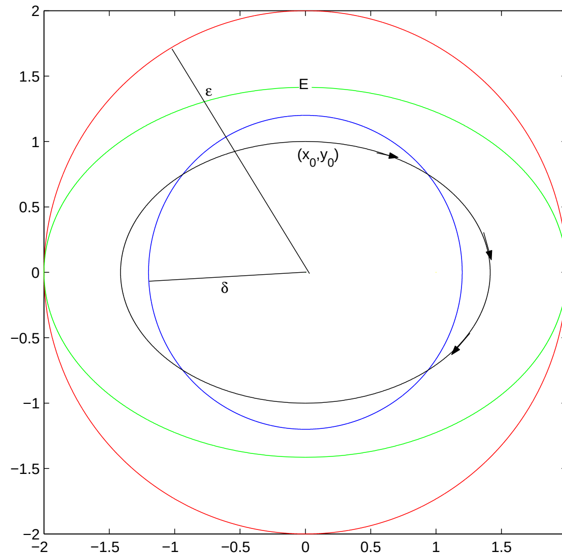
E' immediato notare che $V(x,y) = x^2 + 2y^2$ è un integrale primo di questo sistema. Le curve di livello sono le ellissi

$$x^2 + 2y^2 = c\,.$$

Usiamo questo per mostrare che l'origine è un punto di equilibrio stabile. Si fissi per questo $\epsilon > 0$ e sia c così piccolo che l'ellisse $x^2 + 2y^2 = c$ sia contenuta nel disco di centro l'origine e raggio ϵ. Indichiamo con E quest'ellisse.

Sia $\delta > 0$ così piccolo che la circonferenza di centro l'origine e raggio δ sia contenuta nell'ellisse E, si veda la figura 10.6.

Figura 10.6: Integrali primi e stabilità



Sia (x_0, y_0) un dato iniziale di norma minore di δ. L'orbita che lo contiene è contenuta in un ellisse che appartiene alla regione interna all'ellisse E e quindi in particolare nell'intorno dell'origine di raggio ϵ. Ciò mostra la stabilità. $\blacksquare$

In generale lo studio geometrico delle curve di livello non è facile. Vogliamo dare un risultato di stabilità che usi l'idea dell'esempio precedente, ma che non richieda la determinazione esplicita delle curve di livello. Questo risultato si applica a sistemi della forma

$$mx' = y\,,\quad y' = -\phi(x) \qquad \text{ossia} \qquad mx'' = -\phi(x) \qquad \text{con } \phi(0) = 0. \tag{10.20}$$

L'origine, ossia il punto di coordinate $(0,0)$, è punto d'equilibrio di questo sistema.

Si sa che a questo sistema si può associare sia l'energia potenziale

$$\Phi(x)\,, \qquad \Phi(x) = -\int_0^x \phi(s)\,\mathrm{d}s$$

sia l'"energia" totale

$$E(x,y) = \frac{1}{2}my^2 + \Phi(x)\,.$$

Diciamo che $\Phi(x)$ ha *minimo stretto* in 0 quando

$$\Phi(0) < \Phi(x) \qquad \forall x \neq 0\,.$$

Analoga definizione per l'energia totale $E(x,y)$.

Vale:

Teorema 283 (di Lagrange) *Supponiamo che* 0 *sia uno zero isolato di* $\phi(x)$*. Se* 0 *è punto di minimo stretto per* $\Phi(x)$ *allora il punto d'equilibrio* $(0,0)$ *del sistema (10.20) è stabile.*

Dim. Il fatto che $\Phi(x)$ abbia un minimo stretto in zero implica che l'energia totale $E(x,y)$ ha minimo stretto in $(0,0)$:

$$0 = E(0,0) < E(x,y) \qquad \forall (x,y) \neq (0,0)\,.$$

Inoltre, l'energia totale è una funzione continua. Fissiamo $\epsilon > 0$ e consideriamo il numero

$$E_0 = \min\{E(x,y)\,, \quad ||(x,y)|| = \epsilon\}\,.$$

Essendo $(0,0)$ un minimo stretto, segue che $E_0 > 0$.

Scegliamo ora $\sigma > 0$ tale che se $||(x,y)|| < \sigma$ allora si abbia

$$E(x,y) < E_0/2\,.$$

Sia (x_0, y_0) un dato iniziale con $||(x_0,y_0)|| < \sigma$. Sia $(x(t), y(t))$ la soluzione corrispondente a tale dato iniziale. Lungo questa soluzione l'energia si conserva e quindi

$$E(x(t), y(t)) = E(x_0, y_0) < E_0/2\,.$$

Dunque, $||(x(t), y(t))||$ non può mai prendere il valore ϵ: se ciò avvenisse per un valore $\bar{t}$ di t avremmo infatti $E(x(\bar{t}), y(\bar{t})) > E_0/2$. Quindi, essendo $t \to ||(x(t), y(t))||$ una funzione continua, si ha sempre

$$||(x(t), y(t))|| < \epsilon\,.$$

Ossia, se il dato iniziale ha norma minore di σ, in ogni istante successivo la soluzione ha norma minore di ϵ. L'arbitrarietà di ϵ prova che il punto di equilibrio $(0,0)$ è stabile. ∎

Dalla definizione di $\Phi(x)$ si vede immediatamente:

Corollario 284 *Le condizioni del Teorema 283 sono soddisfatte se la funzione $\phi(x)$ verifica*

$$x\phi(x) > 0\,.$$

Esempio 285 Consideriamo l'equazione

$$x'' = -\sin x$$

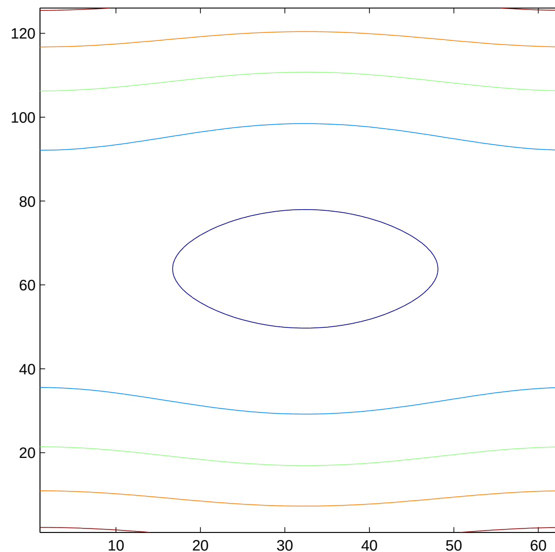
che descrive il moto di un pendolo. Le condizioni del Corollario 284 sono soddisfatte e quindi il punto di equilibrio $(0,0)$ è stabile.

L'energia totale è

$$E(x,y) = \frac{1}{2}y^2 + \cos x\,.$$

Alcune delle sue curve di livello intorno al punto d'equilibrio $(0,0)$ sono riportate nella figura 10.7. Dato che l'energia si conserva, una soluzione che ha dato iniziale all'interno della curva di livello chiusa non esce dalla regione delimitata da questa curva, in accordo con l'asserto del Teorema di Lagrange. Se però il dato iniziale è "lontano" da $(0,0)$ allora la traiettoria del sistema ad esso corrispondente si allontana dal punto di equilibrio. ∎

Figura 10.7: Il caso considerato nell'esempio 285



10.4.2 Stabilità asintotica e perturbazioni

E' importante osservare che il teorema di Lagrange dipende da una proprietà di minimo che non si conserva sotto l'azione di piccole perturbazioni. E in effetti piccole perturbazioni possono distruggere la proprietà di stabilità, come mostrano gli esempi seguenti.

Esempio 286 Consideriamo il sistema scalare

$$x' = \epsilon x\,.$$

Se $\epsilon = 0$ allora il punto d'equilibrio 0 è stabile, mentre è instabile per ogni $\epsilon > 0$.

Questo esempio non verifica le ipotesi del teorema di Lagrange. Per trovare invece un esempio che verifica le ipotesi del teorema di Lagrange, consideriamo

$$x'' = \epsilon x' - x\,.$$

Se $\epsilon = 0$ questa è l'equazione del moto armonico e verifica le ipotesi del Teorema di Lagrange. Dunque, $(0,0)$ è punto d'equilibrio stabile come d'altra parte si verifica anche direttamente notando che la matrice corrispondente ha i due autovalori $\pm i$, distinti e puramente immaginari. Il punto d'equilibrio $(0,0)$ diviene instabile per ogni $\epsilon > 0$. ∎

Esempio 287 Consideriamo ora il sistema

$$x' = -y - x^5\,, \qquad y' = x - y^3\,.$$

Questo sistema non ammette integrali primi. Però, se $(x(t), y(t))$ è una soluzione,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left|\left|\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}\right|\right|^2 = -2x^6(t) - 2y^4(t) < 0\,.$$

Ciò mostra che la norma decresce e che quindi il punto di equilibrio $(0,0)$ è stabile. Si può provare che è addirittura asintoticamente stabile. Però, la "piccola" perturbazione che consiste nel trascurare i termini x^5 ed y^3, che è una perturbazione infinitesima di ordine superiore al primo (rispetto ad (x,y)), ne distrugge la stabilità asintotica. ∎

Invece, la stabilità asintotica dei sistemi lineari si conserva sotto l'azione di "piccole" perturbazioni. Questa è una proprietà importantissima per le applicazioni, ed anche molto comoda in pratica perché dà un test facile da usare. Enunciamo il risultato senza provarlo:

Teorema 288 *Sia*

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}(\mathbf{x}) \tag{10.21}$$

e supponiamo che $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. *Supponiamo inoltre:*

- *la matrice A abbia tutti gli autovalori con parte reale negativa, così che il punto di equilibrio* $\mathbf{0}$ *di*

  $$\mathbf{x}' = A\mathbf{x}$$

  è asintoticamente stabile;

- *Valga*

  $$\lim_{\mathbf{x}\to 0} \frac{\mathbf{f}(\mathbf{x})}{||\mathbf{x}||} = \mathbf{0}\,.$$

Sotto queste condizioni il punto d'equilibrio $\mathbf{0}$ *è asintoticamente stabile per il sistema (10.21).*

Il teorema precedente può anche essere precisato: si può provare che se la matrice A ha un autovalore con parte reale **positiva** allora l'origine **non** è stabile per il sistema (10.21).

Questo teorema in pratica si usa come segue: si ha un sistema della forma

$$\mathbf{x}' = \mathbf{g}(\mathbf{x})\,, \text{ con } \mathbf{g}(\mathbf{0}) = \mathbf{0}\,. \tag{10.22}$$

Dato che $\mathbf{g}(0) = 0$, lo sviluppo di McLaurin di $\mathbf{g}$ arrestato al primo ordine dà

$$\mathbf{g}(\mathbf{x}) = A\mathbf{x} + \mathbf{f}(\mathbf{x})\,, \qquad A = J_{\mathbf{g}}(\mathbf{0})$$

e si sa che $\mathbf{f}(\mathbf{x})$ è infinitesimo di ordine maggiore di 1 rispetto a $||\mathbf{x}||$, ossia

$$\lim_{\mathbf{x}\to 0} \frac{\mathbf{f}(\mathbf{x})}{||\mathbf{x}||} = \mathbf{0}\,.$$

Se gli autovalori di A hanno tutti parte reale negativa, allora il punto d'equilibrio 0 del sistema (10.22) è asintoticamente stabile.
Si usa dire che il sistema lineare

$$\mathbf{x}' = A\mathbf{x}$$

si ottiene linearizzando in 0 il sistema $\mathbf{x}' = \mathbf{g}(\mathbf{x})$.

Quando si usa questo teorema per studiare la stabilità asintotica di un punto di equilibrio, si dice che si studia la *stabilità in prima approssimazione.*

Indice analitico